Linear algebra and geometry

Jakob Scholbach

May 23, 2025

Contents

Chapter 0

Preface

These are lecture notes on *Linear algebra and geometry*, offered in Spring 2023 and subsequent years at the University of Padova to an audience of engineering students.

Chapter 1

The Complex Numbers

In this brief chapter, we survey some fundamental properties of the complex numbers, namely the basic arithmetic of complex numbers, the fundamental theorem of algebra and the complex exponential function.

1.1 Basic Properties

Definition 1.1. The *complex numbers* $\mathbf{C}$ are the set

$$\mathbf{C} = \{(a, b), \text{ with } a, b \in \mathbf{R}\}.$$

Addition and multiplication of complex numbers are defined by

$$(a_1, b_1) + (a_2, b_2) := (a_1 + a_2, b_1 + b_2)$$
$$(a_1, b_1) \cdot (a_2, b_2) := (a_1 a_2 - b_1 b_2, a_1 b_2 + a_2 b_1).$$

We usually write such a pair as

$$a + ib := (a, b).$$

This way of writing a complex number is also referred to as the *algebraic form*. In this notation, we obtain

$$(a_1 + ib_1)(a_2 + ib_2) = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1).$$

In particular, the fundamental equation holds:

$$i^2 (:= ii) = -1.$$

We consider the real numbers $\mathbf{R}$ as the subset $\{(a, 0), a \in \mathbf{R}\} = \{a + i0\} \subset \mathbf{C}$. The *real part* and *imaginary part* are defined as

$$\Re(a + ib) := a, \Im(a + ib) := b.$$

Another customary notation is to write z for complex numbers, i.e., $z = a + ib$.

The sum of complex numbers just amounts to adding the real part and the imaginary parts separately.

Sometimes, we refer to the set of complex numbers also as the complex plane, given that we can switch back and forth between writing $z = a + ib$ and $z = (a, b)$, i.e., a point on the plane specified by its two coordinates.

Regarding the geometric properties of $\mathbf{C}$, we use the following standard terms:

Definition 1.2. The *complex conjugation* is the mapping

$$\overline{?} : \mathbf{C} \to \mathbf{C}, z = a + ib \mapsto \overline{z} := a - ib.$$

The *absolute value* is the mapping

$$|?| : \mathbf{C} \to \mathbf{R}^{\geqslant 0}, z = a + ib \mapsto |z| := \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

Geometrically, the operation of taking the complex conjugation amounts to reflecting the points in the complex plane along the x-axis (also known as real axis). By the Pythagorean theorem, if we depict z as a point in the plane with the coordinates (a, b), then $|z|$ is the distance of that point to the origin. We note that $|z| = 0$ holds exactly if $z = 0$. If $z = a(= a + 0b)$ happens to be a real number, then $|z| = |a|$ is the usual absolute value of a real number: $|a| = a$ if $a \geqslant 0$ and $|a| = -a$ if $a < 0$.

Lemma 1.3. If $z \neq 0$ is a non-zero complex number, i.e., $z = a + ib$ with a or b (or both) being nonzero, then

$$z^{-1} = \frac{1}{z}$$

exists (as a complex number). It can be computed as

$$z^{-1} = \frac{\overline{z}}{z \cdot \overline{z}}.$$

Example 1.4. If $z = 2 + 3i$, we compute $\overline{z} = 2 - 3i$, and $z\overline{z} = 2^2 + 3^2 = 13$. Thus $z^{-1} = (2 - 3i)/13 = \frac{2}{13} - \frac{3}{13}i$.

Proof. Indeed, if we multiply the above expression with z, we get

$$zz^{-1} = z\frac{\overline{z}}{z \cdot \overline{z}} = 1.$$

Note how this formula is an actual simplification, since $z \cdot \overline{z}$ is a real (as opposed to complex) number, so dividing by it is easily done. It is also possible to state the above formula without referring to the complex conjugate, but the formula becomes less transparent:

$$z^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i.$$

Given that we can form the reciprocal of any nonzero complex number, we can also divide any complex number $w \in \mathbf{C}$ like so:

$$\frac{w}{z} = \frac{w\overline{z}}{z\overline{z}} = \frac{w\overline{z}}{|z|^2}.$$

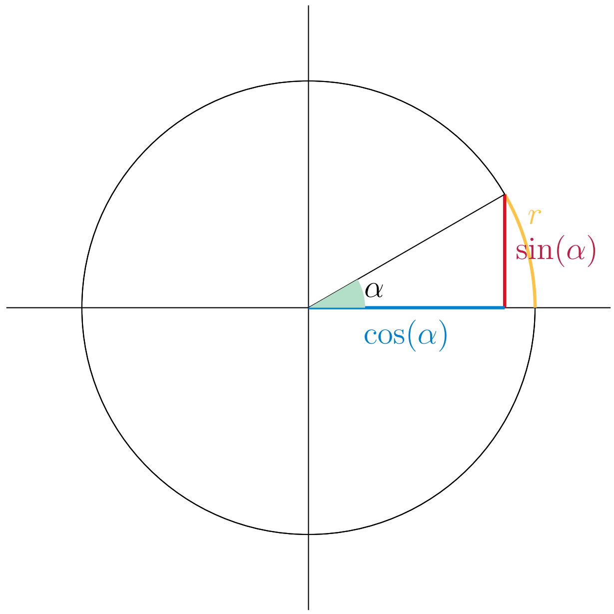
1.1.1 Trigonometric form

The multiplication of complex numbers reveals its essence best by writing complex numbers in a different form, known as the trigonometric form.

For any two complex numbers $z_1 = a_1 + b_1 i, z_2 = a_2 + b_2 i \in \mathbf{C}$, one has

$$\begin{aligned}
|z_1 z_2| &= |a_1 a_2 - b_1 b_2 + (a_1 b_2 + a_2 b_1)i| \\
&= \sqrt{(a_1 a_2 - b_1 b_2)^2 + (a_1 b_2 + a_2 b_1)^2} \\
&= \sqrt{(a_1 a_2)^2 - 2a_1 a_2 b_1 b_2 + (b_1 b_2)^2 + (a_1 b_2)^2 + 2a_1 b_2 a_2 b_1 + (a_2 b_1)^2} \\
&= \sqrt{(a_1^2 + b_1^2)(a_2^2 + b_2^2)} \\
&= |z_1||z_2|.
\end{aligned}$$

In particular, if $z \neq 0$, we have that $\frac{z}{|z|}$ is a complex number with absolute value 1. I.e., its distance to the origin is 1. Yet in other words, the complex number $\frac{z}{|z|}$ lies on the circle (around the origin) with radius 1, which is also known as the *unit circle*.



The angle α between the positive x-axis and the line segment joining the origin and the point $\frac{z}{|z|}$ is called the *argument* of z. It is denoted by $\arg z$, and is commonly measured in radian (not in degree).

It is therefore possible to write

$$z = |z| \cdot (\cos\alpha + i \sin\alpha) \tag{1.5}$$

(or $z = |z| \cdot (\cos(\arg z) + i \sin(\arg z))$). Note that $|z|$ is a real number, $|z| \geqslant 0$. Also the argument α is a real number. It is common to require α to satisfy

$$-\pi < \alpha \leqslant \pi.$$

This choice of the argument is called the *principal argument*. If we impose this requirement, there is a *unique* value of α such that (1.5) holds. However, it may be possible to express z in a similar form, but for a different value of the argument: indeed, adding (or subtracting 2π) to α amounts geometrically to rotating by $2\pi = 360°$ counter-clockwise (or, for subtracting, by 2π but clockwise), which does not affect the resulting point.

More precisely, we have the following fact: an equation

$$r \cdot (\cos\alpha + i \sin\alpha) = s \cdot (\cos\beta + i \sin\beta)$$

holds exactly if

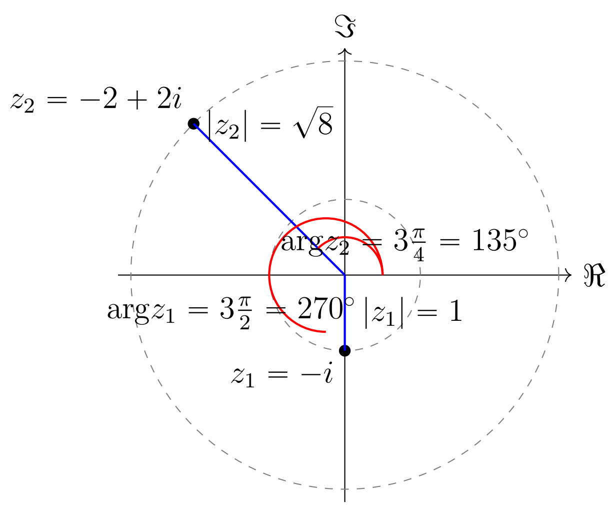
(1) $r = s$ and

(2) $\alpha - \beta$ is an integer multiple of 2π. We also say that α and β are *equal modulo* 2π to express this, and write

$$\alpha \equiv \beta \mod 2\pi. \tag{1.6}$$

This ambiguity of the argument has to taken into account when solving equations involving complex numbers in trigonometric form. See Exercise 1.5 and its solution for a worked example.

Example 1.7.
- $z = 1 = 1 + 0i$ has absolute value $|z| = 1$ and its argument is $\arg z = 0$.
- $z = -i$ has absolute value $|-i| = 1$ and its argument is $3\frac{\pi}{2}$ (which amounts to $270°$).
- $z = -2 + 2i$ has absolute value $|z| = \sqrt{8} = 2\sqrt{2}$. Its (principal) argument is $\arg z = 3\frac{\pi}{4}$.



Lemma 1.8. If $z_1, z_2 \in \mathbf{C}$ are given in trigonometric form, i.e.,

$$z_1 = |z_1|(\cos(\arg z_1) + i \sin(\arg z_1))$$

and likewise for z_2, the product $z_1 z_2$ is given by *multiplying* the absolute values, and *adding* the arguments. That is:

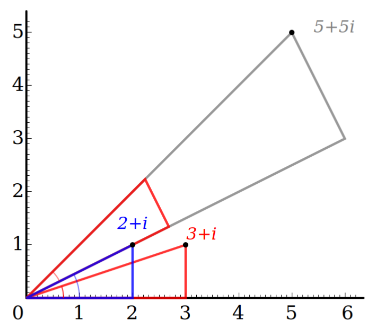
$$z_1 z_2 = |z_1||z_2| \left(\cos\left(\arg z_1 + \arg z_2\right) + i \sin\left(\arg z_1 + \arg z_2\right)\right).$$

Proof. We have already noted that $|z_1 z_2| = |z_1||z_2|$. Concerning the arguments, let us write $\alpha_1 = \arg z_1$ and likewise α_2. Then

$$\begin{aligned}(\cos\alpha_1 + i\sin\alpha_1)(\cos\alpha_2 + i\sin\alpha_2) &= \cos\alpha_1\cos\alpha_2 - \sin\alpha_1\sin\alpha_2 + i\,(\sin\alpha_1\cos\alpha_2 + co \\ &= \cos(\alpha_1+\alpha_2) + i\sin(\alpha_1+\alpha_2).\end{aligned}$$

Here, the first equality is the definition of the multiplication of complex numbers, and the second equality holds by the *angle sum identities*. □

The following illustration is from Wikimedia[1]



1.2 The Fundamental Theorem of Algebra

We have noted above that we can perform addition, subtraction, multiplication and division for complex numbers (division by a non-zero number). These continue to satisfy the same rules that we are familiar with from the real numbers. That is, for any three complex numbers v, w, z we have the following identities:

- $1 \cdot z = z$, $0 + z = z$,
- $z(wv) = (zw)v$, $z + (w + v) = (z + w) + v$ (*associativity* of multiplication and of addition),
- $zw = wz$, $z + w = w + z$ (so-called *commutativity* of multiplication and of addition),
- $z(v + w) = zv + zw$ (*distributivity law*).

[1] By IkamusumeFan - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=42024950.

Each of these identities is checked by unraveling the definitions.

In the parlance of abstract algebra, one says that the complex numbers form a *field*, exactly as the real or the rational numbers do.

Why do we care (in linear algebra) about the complex numbers? The study of eigenvalues, a cornerstone of linear algebra and its applications throughout all sciences, is concerned with finding a zero of a given polynomial. We will apply the following theorem in Corollary 6.10.

Theorem 1.9. (*Fundamental theorem of algebra*) Consider an equation of the form

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0 = 0,$$

where $a_n, \ldots, a_0$ are complex numbers. We assume that $a_n \neq 0$ and $n > 0$ (i.e., that the equation is not just of the form $a_0 = 0$). Then this equation has a complex solution, i.e., there is a complex number z satisfying the above equation.

This is in marked contrast to the real numbers. For example,

$$t^2 + 1 = 0$$

has no real solution since t^2+1 is positive for all real numbers t. If we consider roots within the complex numbers, however, the situation changes: this polynomial has exactly two complex roots, namely i and $-i$, since

$$i^2 = i \cdot i = -1, \text{ and } (-i)^2 = (-i)(-i) = -1.$$

The Fundamental Theorem of Algebra states a much stronger result: not only this polynomial, but every non-constant real polynomial equation, and more generally, every non-constant complex polynomial equation, has a complex solution.

This theorem is famous for a large number of independent proofs:

- There are short proofs only using basic calculus (such as the intermediate value theorem). For example [Oli11].
- There is a beautiful proof whose essence can be understood purely graphically, as is explained for example in this video https://www.youtube.com/watch?v=RBRVL6nP2Dk. (However, to turn that geometric idea into a rigorous proof does require developing more theory.)

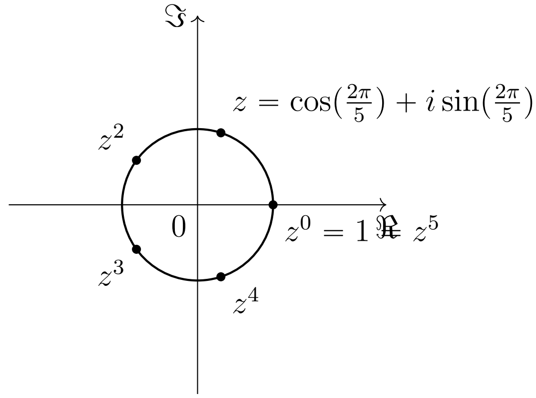
Example 1.10. Let us consider as a special case the equation

$$z^n - 1 = 0.$$

I.e., $z^n = 1$. We are going to solve this equation using the trigonometric form. If $z = r(\cos\alpha + i\sin\alpha)$ we have

$$z^n = r^n(\cos(n\alpha) + i\sin(n\alpha)).$$

This will be equal to $1 = 1 \cdot (\cos 0 + i \sin 0)$ exactly if $r^n = 1$ and if $n\alpha - 0 = n\alpha$ is an integer multiple of 2π, say $n\alpha = 2k\pi$, for an integer $k \in \mathbf{Z}$. Thus, $\alpha = \frac{2k\pi}{n}$.



This is an illustration of the case $n = 5$, in which case there are 5 solutions. (Indeed, $\cos(\frac{6}{5} \cdot 2\pi) + \sin(\frac{6}{5} \cdot 2\pi) = \cos(\frac{1}{5} \cdot 2\pi) + \sin(\frac{1}{5} \cdot 2\pi)$ since winding around by $\frac{6}{5} \cdot 2\pi = 432^\circ$ is the same as winding around by $\frac{1}{5} \cdot 2\pi = 72^\circ$.)

Slightly more generally, equations of the form

$$z^n = w$$

(for a natural number $n \geqslant 1$ and a fixed complex number w) can be conveniently solved as follows. Write $w = r(\cos\alpha + i\sin\alpha)$ in trigonometric form. Then there are n solutions (distinct, unless $w = 0$) of the above equation, namely

$$z = \sqrt[n]{r}(\cos(\frac{\alpha + 2k\pi}{n}) + i \sin\sin(\frac{\alpha + 2k\pi}{n})),$$

where $k \in \{0, 1, \ldots, n-1\}$. This formula is known as *de Moivre's formula*. See Exercise 1.6 for a concrete example.

1.3 The complex exponential function

In elementary real analysis, one defines the (real) exponential, sine and cosine function

$$\exp : \mathbf{R} \to \mathbf{R}$$

by means of the following Taylor series:

$$\begin{aligned} \exp(x) &:= \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots, \\ \sin(x) &:= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{6} + \frac{x^5}{120} \mp \dots, \\ \cos(x) &:= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2} + \frac{x^4}{24} \mp \dots. \end{aligned}$$

One proves there that these three series converge for all real numbers $x \in \mathbf{R}$, and that they define differentiable functions

$$\exp, \sin, \cos : \mathbf{R} \to \mathbf{R}.$$

Moreover, there is the fundamental differential equation

$$\exp' = \exp, \tag{1.11}$$

i.e., the derivative of the exponential function *equals* the exponential function. In addition to this fundamental property, one has the fact that

$$\begin{aligned} \exp(0) &= 1 \\ \exp(x + y) &= \exp x \cdot \exp y. \end{aligned} \tag{1.12}$$

In this section, we briefly deal the corresponding situation in the complex case, which turns out to be very similar and, in a sense, even better.

Definition and Lemma 1.13. Let z be any *complex* number. Then the exact same series as above,

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24} + \dots,$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{6} + \frac{x^5}{120} \mp \dots,$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2} + \frac{x^4}{24} \mp \dots$$

converge, i.e., they are a well-defined complex number. We denote them by $\exp(z)$, $\sin(z)$ and $\cos(z)$, respectively. (Some authors also write e^z for $\exp(z)$).

In addition, the formula (1.12) continues to hold for any two *complex numbers.*

Proof. The proof of the convergence is almost identical to the case of the real exponential, sine and cosine function. Very briefly, the point is that the denominators $n!$ etc. grow fast enough to ensure that the series converge. □

Remark 1.14. The differential equation (1.11) also continues to hold, but we refrain from stating it here in order to avoid a digression about the notion of complex differentiation.

Let us inspect (1.12) in the special case $z = a + ib$. Then

$$\exp z = \exp(a + ib) \overset{(1.12)}{=} \exp a \cdot \exp(ib).$$

By its very definition, the first factor $\exp a$ is just the (usual) real exponential function, so the interesting part is to understand $\exp(ib)$.

Theorem 1.15. For any complex number $z = ib$ we have the fundamental identity

$$\exp(ib) = \cos b + i \sin b.$$

In particular, for $b = \pi$, we have

$$\exp(i\pi) = \cos \pi = -1.$$

Equivalently, we have *Euler's identity*

$$\exp(i\pi) + 1 = 0.$$

Proof. Inserting $z = ib$ into the definition of $\exp(z)$ gives

$$\exp ib = \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} = 1 + ib + \frac{(ib)^2}{2} + \frac{(ib)^3}{6} + \frac{(ib)^4}{24} + \dots$$

We have $(ib)^n = i^n b^n$. We note that i^n only depends on the residue of n after dividing by 4, since $i = i^5 = i^9$ etc. Indeed, $i^4 = i \cdot \dot{i}\dot{i} \cdot i = (-1)(-1) = 1$. We group the terms in the above series according to even and odd n (and basic calculus ensures that this is a legitimate process in this situation):

$$\begin{aligned}
\exp(ib) &= \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} \\
&= \left(1 + \frac{(ib)^2}{2} + \frac{(ib)^4}{24} + \dots\right) + \left((ib) + \frac{(ib)^3}{6} + \frac{(ib)^5}{120} + \dots\right) \\
&= \left(1 - \frac{b^2}{2} + \frac{b^4}{24} \mp\right) + i\left(b - \frac{b^3}{6} + \frac{b^5}{120} \mp \dots\right) \\
&= \cos b + i \sin b. \qquad \square
\end{aligned}$$

1.4 Exercises

Exercise 1.1. Compute the following complex numbers in the form $z = a + ib$:

- $(1-i)(2+i)$
- $(1+i)^3 (:= (1+i)(1+i)(1+i))$
- i^{89}
- $e^{-i\pi}$
- $3(\cos(\pi/2) + i\sin(\pi/2))$
- $\frac{\sqrt{3}+\sqrt{2}i}{\sqrt{2}-\sqrt{3}i}$

Depict these complex numbers on the complex plane.

Exercise 1.2. Prove that a complex number z is a real number exactly if $z = \overline{z}$.

Exercise 1.3. Compute the trigonometric form of the following complex numbers:

- $\sqrt{3} - i$.
- $(1-i)^5$

- $\frac{1+i}{(1-i)(\sqrt{3}+i)}$
- i^n, where n is a natural number.

Exercise 1.4. (Solution at p. 229) Compute the algebraic and the trigonometric form of $z = \left(\frac{i-1}{i+1}\right)^3$.

Exercise 1.5. (Solution at p. 229) Find all complex numbers z satisfying the equation

$$z = 3i|z|\overline{z}.$$

Exercise 1.6. (Solution at p. 230) Compute the solutions of the equation $(\overline{z})^3 = 8i$ in algebraic and in trigonmetric form. Draw a picture of these solutions.

Chapter 2

Systems of linear equations

2.1 Linear equations

Definition 2.1. Let $a, b, c \in \mathbf{R}$ be fixed real numbers. An equation of the form

$$ax + by = c$$

is called a *linear equation* in the *variables* (or *unknowns*) x and y.

More generally, an equation of the form

$$a_1 x_1 + \cdots + a_n x_n = b$$

is called a *linear equation* in the *variables* $x_1, \ldots, x_n$. The real numbers $a_1, \ldots, a_n$ are called the *coefficients* and b is the *constant term* of the equation. The *solution set* consists precisely of those collections (more formally, ordered tuples) of numbers r_1, r_2, up to r_n such that substituting the variables $x_1, \ldots, x_n$ by $r_1, \ldots, r_n$ respectively, the equation holds, i.e., such that

$$a_1 r_1 + \cdots + a_n r_n = b.$$

The name "linear" stems from the geometry of the solution sets, as the following example shows:

Example 2.2. The equation

$$4x + 2y = 3 \tag{2.3}$$

is a linear equation (with coefficients 4 and 2 and constant term 3).

We can solve this equation by subtracting $4x$ from both sides, which gives

$$2y = -4x + 3$$

and dividing by 2, which gives

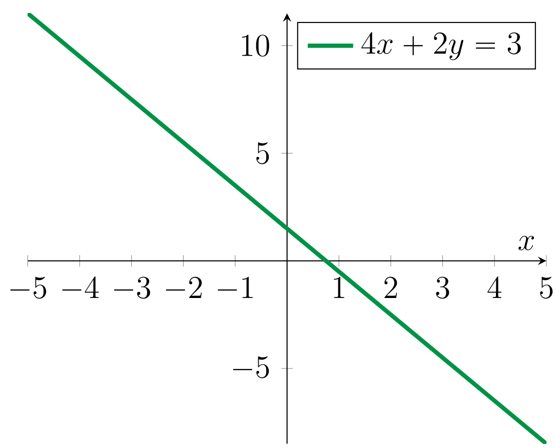
$$y = -2x + \frac{3}{2}. \tag{2.4}$$

In each of these steps, one equation holds precisely if the next one holds. Thus the solution set of (2.3) is the same as the solution set of the equation (2.4). That solution set is therefore the following set:

$$\{(x, -2x + \frac{3}{2}) \text{ with } x \in \mathbf{R}\}.$$

Here we use standard set-theoretic notation, cf. §A. Thus, the above means the set of all pairs $(x, -2x + \frac{3}{2})$, where x is an *arbitrary* real number. In particular, since there are infinitely many real numbers x, this is an infinite set.

Graphically, the solution set is the set of points as depicted below:



In general, any equation of the form

$$ax + by = c$$

with $a \neq 0$ or $b \neq 0$ will have a line as a solution set (what happens if $a = b = 0$?, cf. Exercise 2.6).

Remark 2.5. In the computation above it was critical that we were able to *divide* 3 by 2, i.e., have the rational number $\frac{3}{2}$ at our disposal. The real numbers $\mathbf{R}$ (and also the rational numbers $\mathbf{Q}$) form a so-called *field*, which among other properties means that one can divide by non-zero numbers. Another example of a field are the complex numbers $\mathbf{C}$. The integers $\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ do *not* form a field. Solving linear systems in the integers is somewhat harder than it is in the rationals or reals. This course will focus on discussing linear algebra over the real numbers, with the exception of the discussion of eigenvalues, where the consideration of complex numbers is unavoidable.

Remark 2.6. A great number of equations arising in physics, biology, chemistry and of course mathematics itself are linear. *Nonlinear equations* such as

$$\begin{aligned} x^2 + 4y^3 &= 5 \\ \log(x) - 4\sin(x) &= 0 \end{aligned}$$

are not primarily studied in linear algebra. For such more complicated equations, linear algebra is still useful, however. This is accomplished by replacing such equations by *linear approximations*. The first idea in that direction is the derivative of a function f, which serves as a best linear approximation of a differentiable function. Such linearization techniques are beyond the scope of this lecture.

2.2 Systems of linear equations

Definition 2.7. A *system of linear equations* is a collection of linear equations (involving the same variables). It is also sometimes called a *linear system* or even just a *system*.

The interest in linear systems lies in finding those tuples of numbers satisfying *all* equations at once (as opposed to just one of them, say). We will start with two equations in two variables.

Example 2.8. The equations

$$\begin{aligned} x + y &= 4 \\ x - y &= 1. \end{aligned} \tag{2.9}$$

form a system of linear equations (in the variables x and y).

We solve this system algebraically by subtracting y in the first equation, which gives

$$x = -y + 4,$$

and substituting this into the second equation, which gives

$$(-y + 4) - y = 1,$$

or

$$-2y + 4 = 1$$

or

$$-2y = -3$$

or finally

$$y = \frac{3}{2}.$$

Inserting this back above, gives

$$x = -\frac{3}{2} + 4 = \frac{5}{2}.$$

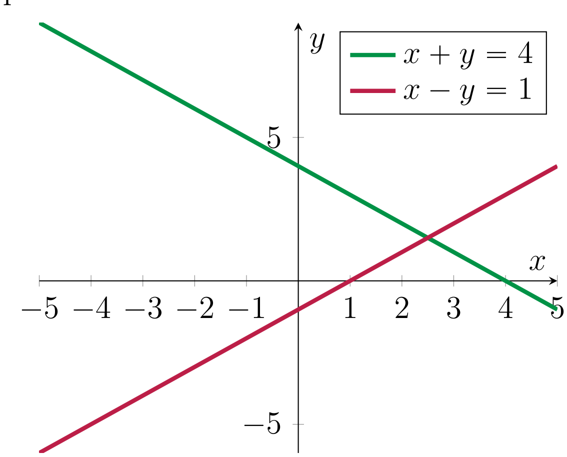
Note that again each equation holds (for given values of x and y) precisely if the preceding one holds. Thus, the original system has the same solution set as the last two equation (together). This system of equations therefore has a *unique* solution, namely

$$(x = \frac{5}{2}, y = \frac{3}{2}).$$

To say the same using different symbols: the solution set of the system (2.9) is a set consisting of a single element:

$$\{(\frac{5}{2}, \frac{3}{2})\}$$

It is very useful to also understand this process geometrically, which we do by plotting the two lines that are the solutions of the individual equations:



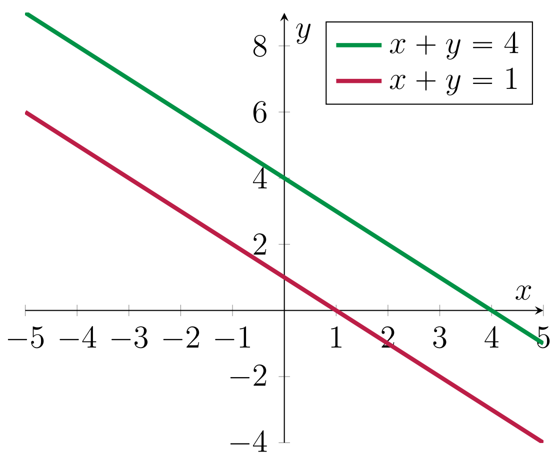
The algebraic computation of having precisely one solution is matched by the fact that two non-parallel lines in the plane (which are the solution sets of the individual equations) exact in precisely one point.

The above linear system (2.9) had exactly one solution. This need not always be the case, as the following examples show:

Example 2.10. The system

$$\begin{aligned} x + y &= 4 \\ x + y &= 1 \end{aligned}$$

has *no solution*. This can be seen algebraically (!) and also geometrically:



The system has no solution, which is paralleled by the fact that two *parallel, but distinct* lines in the plane do not intersect.

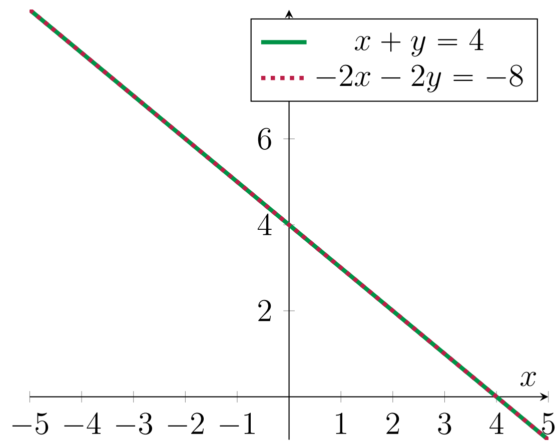
Example 2.11. The system

$$\begin{aligned} x + y &= 4 \\ -2x - 2y &= -8 \end{aligned}$$

has infinitely many solutions, namely all pairs of the form

$$(x, y = 4 - x),$$

with an arbitrary real number x. Geometrically, this is explained by taking the "intersection" of the same line twice.



In other words, even though there are two equations above, they both have the same solution set. Thus, in some sense one of the equations is redundant, i.e., the solution set of the entire system equals the solution set of either of the equations individually.

Summary 2.12. The solution set of an equation of the form

$$ax + by = c$$

is a line (unless both a and b are zero).

The solution set of a system of equations of the form

$$ax + by = c$$

$$dx + ey = f$$

can take three forms:

number of solutions	geometric explanation
exactly one solution	the unique intersection point of two non-parallel lines
no solution	two distinct parallel lines don't intersect
infinitely many solutions	a line intersects itself in infinitely many points

Definition 2.13. A *homogeneous linear system* is one in which the constant terms in all equations are zero. (I.e., in the notation in (2.25) below, $b_1 = \cdots = b_n = 0$.)

Remark 2.14. For a homogeneous linear system, there is always *at least* one solution namely

$$(x_1 = 0, \ldots, x_n = 0).$$

This solution is called the *trivial solution*.

2.3 Elementary operations

The combination of geometric intuition with algebraic computations is very useful. However, the former is of limited use when it comes to systems with three variables, and hardly useful anymore for systems involving four or more variables. We will therefore develop notions and techniques that enable us to handle linear systems more systematically.

Definition 2.15. We say that two linear systems are *equivalent* if they have the same solution sets.

Example 2.16. In Example 2.8, we considered the system

$$\begin{aligned} x + y &= 4 \\ x - y &= 1 \end{aligned}$$

and found that it has a unique solution, namely

$$(x = \frac{5}{2}, y = \frac{3}{2}).$$

Thus the previous system is equivalent to the system

$$\begin{aligned} x &= \frac{5}{2} \\ y &= \frac{3}{2}. \end{aligned}$$

Of course, in comparison to the original system, the latter system is much easier to understand, since one can simply read off the solution without any effort. The purpose of elementary operations is to transform a given system into an equivalent system of which the solutions can be read off.

Definition 2.17. Given a linear system, the following operations are called *elementary operations*:

(1) interchange two equations,

(2) multiply one equation by a *non-zero* (!) number,

(3) add a multiple of one equation to a *different* (!) equation.

These operations are called "elementary" since they are so simple to perform. Their utility comes partly from the following fact:

Theorem 2.18. Consider a linear system. This linear system is equivalent to (i.e., has the same solutions as) any linear system obtained by performing any number of elementary operations.

This theorem, which we will prove later on (Corollary 4.77) when we have more tools at our disposal may sound a little abstract at first sight. It is however actually simple to comprehend and, very importantly, extremely useful in practice.

Example 2.19. Consider the system

$$\begin{aligned} x + 2z &= -1 \\ -2x - 3z &= 1 \\ 2y &= -2. \end{aligned}$$

We add twice the first equation to the second (elementary operation (3)):

$$\begin{aligned} x + 2z &= -1 \\ z &= -1 \\ 2y &= -2. \end{aligned}$$

We interchange the second and third equation (elementary operation (1)):

$$\begin{aligned} x + 2z &= -1 \\ 2y &= -2 \\ z &= -1. \end{aligned}$$

We multiply the second equation by $\frac{1}{2}$ (in other words, we divide it by 2; elementary operation (2)):

$$\begin{aligned} x + 2z &= -1 \\ y &= -1 \\ z &= -1. \end{aligned}$$

We add (-2) times the third equation to the first (elementary operation (3))

$$\begin{aligned} x &= 1 \\ y &= -1 \\ z &= -1. \end{aligned}$$

These steps are combinations of elementary operations. According to Theorem 2.18, the original system is equivalent (i.e., has the same solutions) as the final one. The benefit is, of course, that the solutions of the final system are trivial to comprehend: it has exactly one solution, the triple

$$(x = 1, y = -1, z = -1).$$

Thus, the original system also has exactly that one solution.

2.4 Matrices

It is time to use some better tools to do the bookkeeping needed to solve linear systems. Matrices help doing that. Later on (§4), we will use matrices in a much more profound way.

Definition 2.20. A *matrix* is a rectangular array of numbers. We speak of an $m \times n$-matrix (or m-by-n matrix) if it has m rows and n columns, respectively. If $m = n$, we also call it a *square matrix*.

An $1 \times n$-matrix (i.e., $m = 1$ and n is arbitrary) is called a *row vector*. Similarly, an $m \times 1$-matrix is called a *column vector*.

Example 2.21. It is customary to denote matrices by capital letters. For example,

$$A = \begin{pmatrix} 3 & 4 \\ 0 & -7 \end{pmatrix}$$

is a 2×2-matrix (or square matrix of size 2).

$$B = \begin{pmatrix} 1 & -2 & 0 \\ 1 & 0 & 3 \end{pmatrix}$$

is a 2×3-matrix and

$$C = \begin{pmatrix} 1 & -2 \\ 0 & 1 \\ 0 & -3 \end{pmatrix}$$

is a 3×2-matrix.

The entries of a matrix may also be variables. For example

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is a column vector (or a 2×1-matrix), whose entries are two variables; $\begin{pmatrix} x_1 & x_2 \end{pmatrix}$ is a row vector (or a 1×2-matrix).

Notation 2.22. A matrix whose entries are unspecified numbers is denoted like so:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

Thus, the number a_{ij} is the entry in the i-th row and the j-th column. A more compressed notation expressing the same is

$$A = (a_{ij})_{i=1,\dots,m, j=1,\dots,n}$$

or even just

$$A = (a_{ij}).$$

Definition 2.23. Let

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 (2.24) \qquad (2.25) \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

be a linear system (consisting of m equations, in the unknowns $x_1, \dots, x_n$; the numbers a_{ij} are the coefficients, the numbers $b_1, \dots, b_n$ are the constants).

The *matrix associated to this system* is the following $m \times (n+1)$-matrix (the vertical bar is just there to remind ourselves that the last column corresponds to the constants in the equations above; one also speaks of an *augmented matrix*)

$$A = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right). \tag{2.26}$$

In other words, the matrix is the rectangular array containing the coefficients and the constants of the individual equations, and suppresses the mention of the variables.

Example 2.27. The matrix associated to the system

$$\begin{aligned} x + y &= 4 \\ x - y &= 1 \end{aligned}$$

is the 2×3-matrix

$$\left(\begin{array}{cc|c} 1 & 1 & 4 \\ 1 & -1 & 1 \end{array} \right).$$

Of course, the process of associating a matrix to a linear system can be reversed since any $m \times (n+1)$-matrix gives rise to a linear system: the matrix (2.26) gives rise to the linear system (2.25). For example, the 2×3- matrix

$$\left(\begin{array}{cc|c} 1 & -2 & 0 \\ 1 & 0 & 3 \end{array} \right)$$

gives rise to the linear sytem

$$\begin{aligned} x - 2y &= 0 \\ x &= 3. \end{aligned}$$

2.5 Gaussian elimination

Theorem 2.18 is a useful insight, but it lacks an important feature: it does not directly instruct us how to simplify any given linear system.

(In Example 2.19, we did end up with a particularly simple linear system, but we did not have any a priori guarantee for this to happen. If we had chosen some "stupid" elementary operations instead, we would not have simplified the system.) *Gaussian elimination* is an algorithmic process that does just that: it is a simple procedure that is guaranteed to yield the simplest possible equivalent form of any given linear system.

In view of the correspondence between linear systems and matrices, we will first phrase this process in terms of matrices, and then translate it back to the problem of solving linear systems.

Among the myriad of all possible matrices, the following matrices are the "nice guys".

Definition 2.28. A matrix is in *row-echelon form* (or is called a *row-echelon matrix*) if the following conditions are satisfied:

(1) If there are any zero rows (i.e., consisting only of zeros), then these are at the bottom of the matrix.

(2) In a non-zero row, the first non-zero (starting from the left) is a 1. (It is called the *leading 1*.)

(3) Each leading 1 is to the right of all the leading 1's in the rows above it.

If, *in addition* to the above conditions, each leading 1 is the *only* non-zero entry in its column, then the matrix is in *reduced row-echelon form*.

Maybe saying more than a thousand words is this picture of a row-echelon form. Here, the asterisks indicate an arbitrary number. If, in addition, all the underlined asterisks are zero, then the matrix is a reduced row-echelon matrix.

$$\begin{pmatrix} 0 & 1 & * & \underline{*} & \underline{*} & * & \underline{*} \\ 0 & 0 & 0 & 1 & \underline{*} & * & \underline{*} \\ 0 & 0 & 0 & 0 & 1 & * & \underline{*} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In view of the correspondence between linear systems and matrices, we transport the language of elementary operations (Definition 2.17) to matrices as follows.

Definition 2.29. The following operations on a given matrix A are called *elementary row operations*:

(1) interchange any two rows,

(2) multiply any row by a *non-zero* (!) number,

(3) add a multiple of any row to a *different* (!) row.

Method 2.30. (*Gaussian algorithm* or *Gaussian eliminiation*) Every matrix can be brought to reduced row-echelon form by a sequence of elementary row operations. This can be achieved using the following algorithmic process:

(1) If the matrix consists only of zeros, stop: then the matrix is in reduced row-echelon form.

(2) Otherwise, find the first column from the left having a non-zero entry. Call this entry a. Interchange this row (elementary operation (1)) so that it is in the top position.

(3) Multiply the new top row by $\frac{1}{a}$ (elementary operation (2); note this is possible since $a \neq 0$, see also Remark 2.5). Thus the first row has a leading 1.

(4) By adding appropriate multiples of the first row to the remaining rows (elementary operation (3)), ensure that the entry between the leading 1 are all zero.

From this point on, the first row is not touched anymore, and the four steps above are applied to the matrix consisting of the remaining rows.

This produces a (possibly not reduced) row-echelon form. It can be finally brought into reduced row-echelon form by adding appropriate multiples of rows with leading 1's to the rows above them (elementary operation (3)), beginning at the bottom.

Example 2.31. We apply the Gaussian algorithm to the matrix

$$\begin{pmatrix} 1 & 2 & 5 & 7 \\ 2 & 1 & 4 & 2 \\ 5 & 4 & 13 & 11 \end{pmatrix}.$$

The first three steps don't change the matrix (since the top-left entry is already 1). Step (4): add -2 times the first row to the second row; and add −5 times the first row to the third, which gives

$$\begin{pmatrix} 1 & 2 & 5 & 7 \\ 0 & \underline{-3} & -6 & -12 \\ 0 & -6 & -12 & -24 \end{pmatrix}$$

The remaining steps only affect the second and third row. Step (2) picks the second row, and $a = -3$ (underlined). It is already in the top position (the first row being discarded for the remainder of the algorithm), so Step (2) does not change the matrix. Step (3) gives the matrix

$$\begin{pmatrix} 1 & 2 & 5 & 7 \\ 0 & 1 & 2 & 4 \\ 0 & -6 & -12 & -24 \end{pmatrix}$$

Step (4) adds −6 times the second row to the third, which gives

$$\begin{pmatrix} 1 & 2 & 5 & 7 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

At this point also the second row is discarded, which leaves only the last row, which consists of zeros. By Step (1), the algorithm stops at this point.

This matrix is in row-echelon form, but not yet reduced. To reduce it, add −2 times the second row to the first, which gives

$$\begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

When applied to matrices associated to linear systems, Gaussian eliminiation becomes very useful for solving linear systems:

Method 2.32. (1) Form the augmented matrix corresponding to the given linear system.

(2) Perform Gaussian elimination to that matrix (Method 2.30), giving a reduced row-echelon matrix.

(3) If a row of the form

$$0\ 0\ \ldots\ 0\ 1$$

occurs, the system has *no* solutions.

(4) Otherwise, we call the variables corresponding to the columns *not* containing a leading 1 *free variables*. The values of these variables can be chosen to be *arbitrary* real numbers. The variables that correspond to columns that do contain a leading 1 are uniquely specified by these free variables. Their values can be determined by solving the equations corresponding to the row-echelon matrix for the leading variables.

Example 2.33. Consider the system

$$\begin{aligned} x + 2y + 5z &= 7 \\ 2x + y + 4z &= 2 \\ 5x + 4y + 13z &= 11. \end{aligned}$$

The augmented matrix associated to this is the one in Example 2.31. Gaussian algorithm brings it into the reduced row-echelon form

$$\begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This matrix corresponds to the linear system

$$\begin{aligned} x + z &= -1 \\ y + 2z &= 4 \\ 0 &= 0. \end{aligned}$$

According to Theorem 2.18, this linear system has the same solution set as the original one.

The leading variables are x and y, so that z is a free variable. Solving the second equation for y then gives

$$y = 4 - 2z,$$

and similarly

$$x = -1 - z.$$

We obtain that the solution set to the original linear system consists of triples of the form

$$(x = -1 - z, y = 4 - 2z, z),$$

in which z is an arbitrary number (and x and y are determined by z as indicated).

2.6 Exercises

Exercise 2.1. Describe all the solutions of the equation

$$x + y = 3.$$

Draw a picture of that solution set. Is it a homogeneous equation?

Exercise 2.2. Consider the equation

$$x = 3.$$

What is its solution set?

Consider the same equation, but now with two variables x and y being present (so we could rewrite the equation as $x + 0 \cdot y = 3$ in order to emphasize the presence of y). What is the solution set this time?

Exercise 2.3. Consider the system

$$\begin{aligned} 2x_1 - x_2 + x_3 + x_4 &= 1 \\ 5x_2 - 3x_3 - 5x_4 &= -3 \\ 3x_1 - 4x_2 + 3x_3 + 4x_4 &= 3. \end{aligned}$$

What is the matrix associated to that system? Using Method 2.32, find all solutions of that system.

Exercise 2.4. Consider the (augmented) matrix

$$A := \left(\begin{array}{cccccc|c} 1 & 0 & 3 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

What type of matrix is that? (I.e., what $m \times n$-matrix.) If $A = (a_{ij})$, what is a_{13} and a_{31}? What is the linear system associated to that matrix? (Hint: one equation reads "$\cdots = 3$". For consistency, call the variables $x_1, x_2, \ldots, x_6$.)

Is the matrix in row-echelon form? Is it in reduced row-echelon form? If not, use the Gaussian algorithm (Method 2.30) in order to transform it into reduced row-echelon form. Name the columns which contain a leading 1 (Hint: there are 4 of them). Which variables are free, which variables are not free? Use Method 2.32 and solve the linear system associated to that augmented matrix.

Exercise 2.5. Using Method 2.32, find all solutions of the following systems

$$\begin{aligned} x + y - z &= 1 \\ 3x - y + 2z &= 5 \\ 4x + z &= 6. \end{aligned}$$

and

$$\begin{aligned} x + y - z &= 1 \\ 3x - y + 2z &= 0 \\ x + y - 2z &= 2. \end{aligned}$$

Exercise 2.6. (Solution at p. 231) Let

$$ax + by = c$$

be a linear equation. For which values of a, b and c does this equation have no solution? For which values of a, b and c does it have infinitely many solutions?

Exercise 2.7. Compute the reduced row-echelon form of the matrices associated to the linear systems in Example 2.8, Example 2.10 and Example 2.11.

Exercise 2.8. Consider the system

$$\begin{aligned} x + y &= 1 \\ x - y &= b, \end{aligned}$$

where b is a real number. What is its solution set? Illustrate the system geometrically for $b = 0$ and for $b = 1$.

Exercise 2.9. Consider the system

$$\begin{aligned} ax + by &= 1 \\ x - y &= 2. \end{aligned}$$

Here x and y are the variables and a and b are the coefficients.

(1) For which values of a and b does the system above have no solution?

(2) For which values does it have exactly one solution?

(3) For which values does it have infinitely many solutions?
Explain your findings algebraically and geometrically.

Exercise 2.10. (Solution at p. 231) Find the solutions of the system

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 0 \\ -2x_1 - 3x_2 + x_3 &= 1 \\ x_2 - x_3 &= 1. \end{aligned}$$

Exercise 2.11. The linear system in the variables x_1, x_2, x_3, x_4 associated to the matrix

$$\left(\begin{array}{cccc|c} 2 & -1 & 1 & -1 & 1 \\ 0 & 1 & -3 & 1 & 3 \\ 2 & 1 & -4 & 1 & 6 \\ 2 & 0 & -2 & 1 & 2 \end{array}\right)$$

has only one solution. Find it!

Exercise 2.12. (Solution at p. 232) Find the solutions of the following linear system in the variables $x_1, \dots, x_4$:

$$\begin{aligned} x_1 - x_2 + x_3 &= -2 \\ x_3 - x_4 &= 1 \\ x_1 - x_2 + x_4 &= -3 \\ x_1 - x_2 + 3x_3 - 2x_4 &= 0. \end{aligned}$$

Exercise 2.13. Solve the following linear system, where h is a parameter, and x, y are the unknowns:

$$\begin{aligned} x + hy &= 4 \\ 3x + 6y &= 8. \end{aligned}$$

For selected values of h, illustrate the solution set graphically.

Exercise 2.14. (Solution at p. 233) Solve the following linear system, where h is a parameter and x, y, z are the unknowns:

$$\begin{aligned} (4-h)x - 2y - z &= 1 \\ -2x + (1-h)y + 2z &= 2 \\ -x + 2y + (4-h)z &= 1. \end{aligned}$$

Exercise 2.15. For any $t \in \mathbf{R}$ consider the homogeneous linear system associated to the matrix

$$\left(\begin{array}{cccc|c} 2 & 0 & 1 & -t & 0 \\ 1 & -2 & 0 & 3 & 0 \\ 4 & -4 & t & 5 & 0 \end{array}\right).$$

(1) Solve the system for $t = 0$.

(2) Solve the system for all $t \in \mathbf{R}$.

Exercise 2.16. Solve the system

$$\begin{aligned} x_1 - x_3 + 2x_4 &= 0 \\ x_2 + 2x_3 - 2x_4 &= 0 \\ x_1 + x_2 + x_3 &= 0. \end{aligned}$$

Exercise 2.17. (Solution at p. 233) Consider the linear system (in the unknowns x_1, x_2, x_3):

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ x_1 - x_3 &= 0. \end{aligned}$$

Is there any $t \in \mathbf{R}$ such that $(1 - t, 2 + 3t, 4t)$ is a solution of that system?

Exercise 2.18. Consider the following linear system (in the unknowns x_1, x_2, x_3):

$$\begin{aligned} x_1 - x_2 + 3x_3 &= 0 \\ x_1 - x_2 &= 1. \end{aligned}$$

Show that there is exactly one $t \in \mathbf{R}$ such that the vector $(3 + t, 2 + t, \frac{2}{3} + t)$ is a solution of that system.

Exercise 2.19. (Solution at p. 234) Do there exist $q, t \in \mathbf{R}$ such that the vector

$$(x_1, x_2, x_3) = (1 + t, t + q, -t + 2q + 1)$$

satisfies

$$3x_1 + 2x_2 - x_3 = 5?$$

Exercise 2.20. (Solution at p. 234) Find a polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

such that $p(1) = 0$ and $p(2) = 3$. Is there a unique such polynomial?

Exercise 2.21. Find the solution of the linear system associated to the following augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & 1 & 2 & 3 & 1 \\ 2 & 0 & 1 & 2 & 1 \\ 1 & 3 & 5 & 7 & 2 \end{array}\right).$$

Exercise 2.22. For any $\alpha \in \mathbf{R}$ find the solutions of the system associated to the matrix

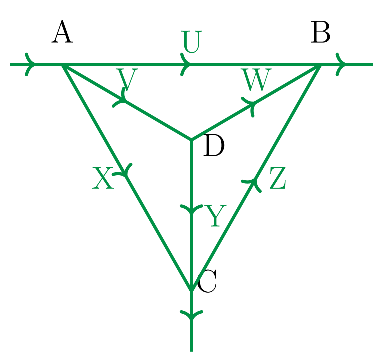
$$\left(\begin{array}{cccc|c} 1 & 1 & 2 & 3 & -1 \\ 2 & 0 & 1 & 2 & \alpha \\ 1 & 3 & 5 & 7 & 0 \end{array}\right).$$

Exercise 2.23. Consider the following linear system in the unknowns x, y, z, which depends on the parameter $\alpha \in \mathbf{R}$:

$$\begin{aligned} 2x - y + z &= 1 \\ (\alpha + 2)x - 2y + \alpha z &= -\alpha. \end{aligned}$$

Determine the solution set of this system for each value of α.

Exercise 2.24. The following extended exercise showcases the usage of linear algebra in network analysis. An idealized city consists of the following streets U to Z, with four intersection points A to D. The streets are all one-way streets:



At the point labelled A, 500 cars per hour drive into the city, and at B, 400 cars exit the city, while at C 100 cars exit the city per hour.

Describe the possible scenarios regarding the numbers of cars driving through the streets U, V, W, X, Y and Z.

Chapter 3

Vector spaces

3.1 $\mathbf{R}^2$, $\mathbf{R}^3$ and $\mathbf{R}^n$

Definition 3.1. For $n \geqslant 1$, an *ordered n-tuple of real numbers* is a collection of n real numbers in a fixed order. For $n = 2$, an ordered 2-tuple is usually called an *ordered pair*, and an ordered 3-tuple is called an *ordered triple*. If these numbers are $r_1, r_2, \ldots, r_n$, then the ordered n-tuple consisting of these numbers is denoted

$$(r_1, r_2, \ldots, r_n).$$

For example, $(2, 3)$ is an (ordered) pair. This pair is different from the (ordered) pair $(3, 2)$. It makes good sense to insist on the ordering, e.g., if a pair consists of the information

(“weight of a parcel (in kg)”, “prize (in €)”),

then $(3, 10)$ is of course different from $(10, 3)$. $(\frac{3}{4}, \sqrt{2}, -7)$, $(0, 0, 0)$ are examples of (ordered) 3-tuples. An ordered 1-tuple is simply a single real number.

Definition 3.2. For $n \geqslant 1$, the set $\mathbf{R}^n$ is the set of all ordered n-tuples of real numbers. Thus (see §A for general mathematical notation)

$$\mathbf{R}^n = \{(r_1, r_2, \ldots, r_n) \mid r_1, r_2, \ldots, r_n \in \mathbf{R}\}.$$

Thus, $\mathbf{R}^1 = \mathbf{R}$ is just the set of real numbers. Next, $\mathbf{R}^2$ is the set of ordered pairs of real numbers:

$$\mathbf{R}^2 = \{(r_1, r_2) \mid r_1, r_2 \in \mathbf{R}\}.$$

Of course, here r_1, r_2 are just symbols which have no meaning in themselves, so we can also write

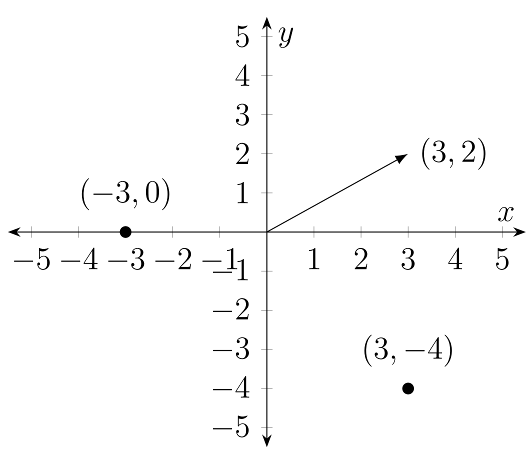
$$\begin{aligned}\mathbf{R}^2 &= \{(x_1, x_2) \mid x_1, x_2 \in \mathbf{R}\}\\ &= \{(x, y) \mid x, y \in \mathbf{R}\}.\end{aligned}$$

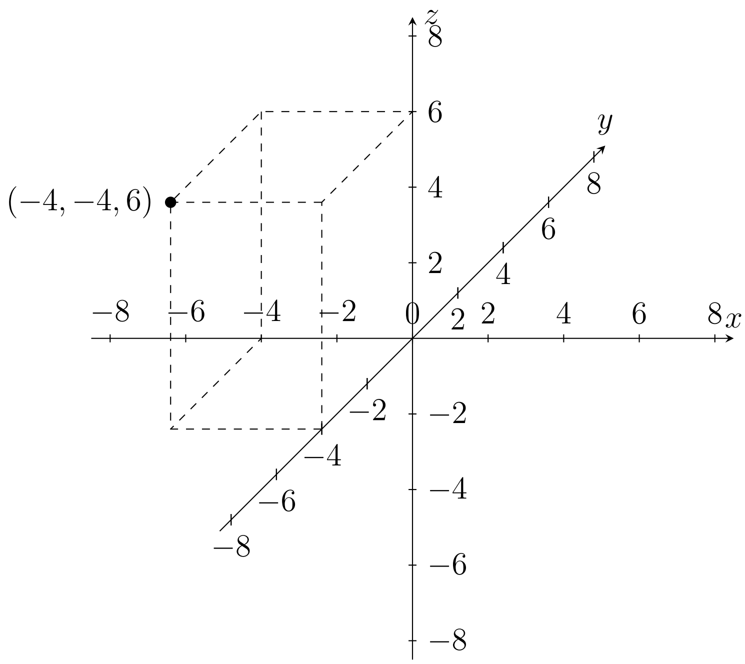
The elements in $\mathbf{R}^n$ are also referred to as *vectors*. Thus, a vector is nothing but an ordered n-tuple. The element $(0, 0, \dots, 0) \in \mathbf{R}^n$ is called the *zero vector*. Instead of writing $(x_1, \dots, x_n)$ we also abbreviate this as x, so that the expression

$$x \in \mathbf{R}^n$$

means that x is an (ordered) n-tuple consisting of n real numbers $x_1, \dots, x_n$. The numbers x_1 etc. are called the *components* of the vector x.

Vectors in $\mathbf{R}$, $\mathbf{R}^2$ and $\mathbf{R}^3$ can be visualized nicely as points on the real line, as points in the plane or as points in 3-dimensional space. It is also common to decorate vectors with an arrow, with the idea of representing a movement or relocation to that point, or in physics a force with a certain strength in a certain direction.





This visualization also helps explain some of the fundamental features of $\mathbf{R}^n$.

3.1.1 Addition of vectors

Definition 3.3. Given two vectors (in the same $\mathbf{R}^n$, i.e., having the same number of components)

$$x = (x_1, \ldots, x_n) \text{ and } y = (y_1, \ldots, y_n) \in \mathbf{R}^n,$$

their *sum* is the vector

$$(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Example 3.4. What is the sum of $(1, 1)$ and $(-2, 1)$? Visualize that sum graphically!

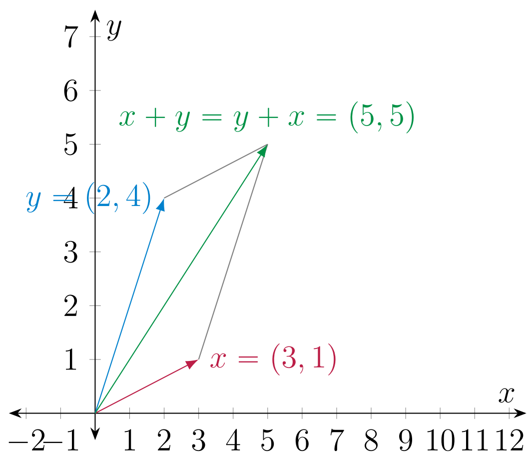
Remark 3.5. The sum of two vectors is only defined if they belong to the *same* $\mathbf{R}^n$: a sum such as $(1, 2) + (3, 4, 5)$ is undefined, i.e. is a meaningless expression.

The sum of vectors has the following crucial properties:

Lemma 3.6. For $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n) \in \mathbf{R}^n$ the following rules hold:

- $x + y = y + x$ (*commutativity of addition*)
- $x + 0 = x$ (adding the zero vector does not change the vector in question)
- $x + (y + z) = (x + y) + z$ (*associativity of addition*)

These identities are easy to prove since they quickly boil down to similar identities for the sum of real numbers. Here is a visual intuition for the commutativity of addition, which is also called the *parallelogram law*.



3.1.2 Scalar multiplication of vectors

Definition 3.7. Given a vector $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ and a real number $r \in \mathbf{R}$, the *scalar multiplication* of x by r is the vector

$$r \cdot x := (r \cdot x_1, \ldots, r \cdot x_n).$$

I.e., every component of x gets multiplied by the number r. Often one just writes rx instead of $r \cdot x$.

Geometrically, the scalar multiplication corresponds to stretching the vector x by the factor r (i.e., if $r > 1$ it is stretching, for $0 < r < 1$ it compresses the vector, for $r < 0$ it additionally flips the direction of the vector).

Example 3.8. What is $4\cdot(-1,3)$? What is $(-\frac{1}{4})\cdot(-1,3)$? Visualize the vector $(-1,3)$ and these results graphically!

Note that in contrast to the sum of vectors the scalar multiplication combines two different entities: a real number and a vector. The scalar multiplication has the following key properties:

Lemma 3.9. For two real numbers $r, s \in \mathbf{R}$ and two vectors $x, y \in \mathbf{R}^n$, the following identities hold:
(1) $r(x+y) = rx + ry$ (*distributivity law*)
(2) $(r+s)x = rx + sy$ (*distributivity law*)
(3) $(rs)x = r(sx)$ (scalar multiplication with a product rs of two real numbers can be computed by first multiplying with s and then with r)
(4) $1x = x$ (scalar multiplication by 1 does not change the vector)
(5) $0x = 0$ (scalar multiplication by 0 gives the zero vector)

Again, these identities are easy to check using that the same rules hold if x, y were just real numbers.

3.1.3 Definition of vector spaces

Definition 3.10. A *vector space* is a set V that is equipped with two functions called the *sum* and the *scalar multiplication*:

$$+ : V \times V \to V, (v,w) \mapsto v+w,$$
$$\cdot : \mathbf{R} \times V \to V, (r,v) \mapsto rv \text{ (or } r\cdot v)$$

satisfying the following conditions. Below $r, s \in \mathbf{R}$ are arbitrary real numbers and $u, v, w \in V$ arbitrary elements of V (also referred to as *vectors*):
(1) $v + w = w + v$ (*commutativity of addition*),
(2) $u + (v + w) = (u + v) + w$,
(3) there is a vector $0 \in V$, called the *zero vector*, such that $0+v = v$ for all $v \in V$,
(4) $r(v + w) = rv + rw$ (*distributive law*),
(5) $(r + s)v = rv + sv$,
(6) $(rs)v = r(sv)$,

(7) $1v = v$,

(8) $0v = 0$ (at the left 0 denotes the real number zero, at the right it denotes the zero vector)

Example 3.11. The sets $\mathbf{R} = \mathbf{R}^1$, $\mathbf{R}^2$, and in general $\mathbf{R}^n$ are vector spaces (where the function $+$ is given by vector addition and $\cdot$ is scalar multiplication). Indeed, the conditions in Definition 3.10 are precisely the properties of vector addition and scalar multiplication noted before in Lemma 3.6 and Lemma 3.9.

Remark 3.12. Recall from §A that the notation appearing in

$$+ : V \times V \to V, (v, w) \mapsto v + w$$

means that $+$ is a function that takes as an input two elements in V, which here are denoted v and w, and produces as an output another element in V. That element is denoted $v + w$. Likewise

$$\cdot : \mathbf{R} \times V, (r, v) \mapsto rv \text{ (or } r \cdot v)$$

means that $\cdot$ is a function whose input is a pair consisting of a real number, here denoted r, and an element in V, and produces as an output an element in V that is denoted rv or $r \cdot v$.

Some authors distinguish notationally between vectors and numbers by writing $\vec{v}$ for vectors and r for numbers. In these notes, we usually do not use that convention.

Example 3.13. The following subsets of $\mathbf{R}^n$ are *not vector spaces.* In each case, draw the set and point out precisely which of the above condition(s) fails.

- $\{(x_1, x_2) \in \mathbf{R}^2 \text{ with } x_1 \geqslant 0\}$,
- $\{(x_1, x_2) \in \mathbf{R}^2 \text{ with } x_1 \neq 0\}$,
- The solution set of the equation
$$3x_1 + 2x_2 = 3.$$
- $\{(x_1, x_2) \in \mathbf{R}^2 \text{ with } x_1 = 0 \text{ or } x_2 = 0\}$.

3.2 Solution sets of homogeneous linear systems

Recall from Definition 2.13 that a *homogeneous linear system* is one on which the constant terms are all zero, i.e., one of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{3.14}$$

In this section, we will see that the solution sets to homogeneous linear systems are vector spaces, which is an extremely important class of examples. We begin by looking at homogeneous linear equations, i.e., a linear system consisting of a single (homogeneous) equation.

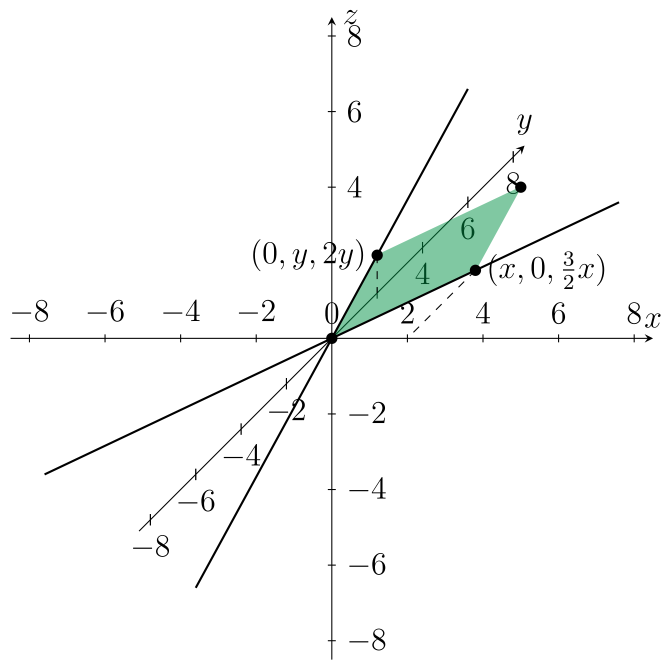
Example 3.15. The homogeneous linear equation

$$3x + 4y - 2z = 0$$

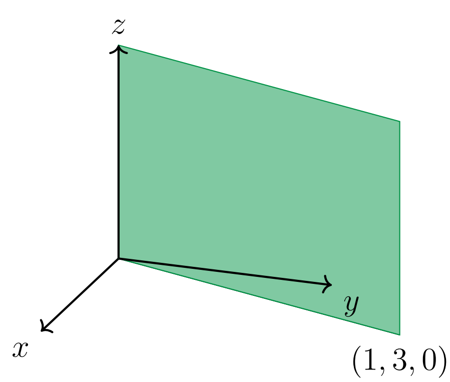
has the solution set

$$\{(x, y, \frac{3x + 4y}{2}) \mid x, y \in \mathbf{R}\}.$$

Indeed, a triple (x, y, z) is a solution to the equation above precisely if $z = \frac{3x+4y}{2}$, and x and y can be arbitrary real numbers. A few concrete elements in this solution set, drawn below, are the points $(0, 0, 0)$, $(2, 0, 3)$, $(0, 1, 2)$. Slightly more generally, triples of the form $(0, y, 2y)$ and $(x, 0, \frac{3}{2}x)$, for arbitrary y, resp. x, are elements in the solution set. These lines (which lie in the $y-z$-plane, resp. in the $x-y$-plane) are also drawn below. Of course, the solution set contains further elements such as the point $(2, 1, 5)$. The green shape is meant to illustrate further elements of the solution set, but of course this is not bounded by the lines in the illustration, instead it stretches out in all directions.



Example 3.16. What equation (in the three variables x, y and z) has the following solution set? Again the picture only shows the solution set partly, it is meant to be extended to the left and below.



We note that both equations have a solution set which is a plane passing through the *origin*, i.e., the point $(0, 0, 0)$. We will want to articulate that this plane is a vector space that lies inside the larger ambient vector space $\mathbf{R}^3$.

Definition 3.17. A *subspace* (or sub-vector space, or vector subspace) V of $\mathbf{R}^n$ is a subset that is – in its own right – a vector space. I.e.,

(1) it contains the zero vector,

(2) for *all* vectors $v, w \in V$, the sum $v + w$ is an element of V, and

(3) for all $v \in V$ and *all* real numbers $r \in \mathbf{R}$, the scalar multiple $r \cdot v$ is required to be an element of V.

More generally, a subset V of another vector space W is a subspace if V satisfies the three preceeding conditions.

We have seen in Example 3.13 a number of sub*sets* of $\mathbf{R}^2$ that fail to be sub*spaces*. In particular, the solution set of the equation $3x_1 + 2x_2 = 3$ is not a vector space since the zero vector $(0, 0)$ is not a solution of this equation. This is not a homogeneous equation (the constant term is 3, but not 0). The next proposition tells us that this is the cause of the failure:

Proposition 3.18. Consider a *homogeneous* linear system in n variables $x_1, \dots, x_n$, and m equations, as in (3.14). Its solution set is a subspace of $\mathbf{R}^n$.

Proof. Let us call S the solution set of the system. I.e., an element $x = (x_1, \dots, x_n)$ belongs to S precisely if it is a solution of the linear system (3.14).

We check the three conditions in Definition 3.17:

- $(0, \dots, 0) \in S$, i.e. the zero vector in $\mathbf{R}^n$ is a solution. Indeed, plugging in zero in all the x_i gives $0 = 0$ for all the m equations, which holds.

- Let $v = (v_1, \dots, v_n)$ and $w = (w_1, \dots, w_n)$ be elements of S. We need to check that $v + w$ is also in S. Recall from Definition 3.3 that $v + w = (v_1 + w_1, \dots, v_n + w_n)$. The m equations of the linear system read

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0,$$

where $i = 1, \dots, m$. Inserting $v_1 + w_1$ for x_1 etc., we get

$$\begin{aligned}
& a_{i1}(v_1 + w_1) + a_{i2}(v_2 + w_2) + \cdots + a_{in}(v_n + w_n) \\
=& a_{i1}v_1 + a_{i1}w_1 + a_{i2}v_2 + a_{i2}w_2 + \cdots + a_{in}v_n + a_{in}w_n \\
=& \underbrace{a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n}_{=0} + \underbrace{a_{i1}w_1 + a_{i2}w_2 + \cdots + a_{in}w_n}_{=0} \\
=& 0 + 0 \\
=& 0.
\end{aligned}$$

This shows that $v + w \in S$.

- In a similar manner, one shows (do it!) that for any $r \in \mathbf{R}$ and $v = (v_1, \dots, v_n) \in S$ the scalar multiple $rv = (rv_1, \dots, rv_n)$ is again in S. □

3.3 Intersection of subspaces

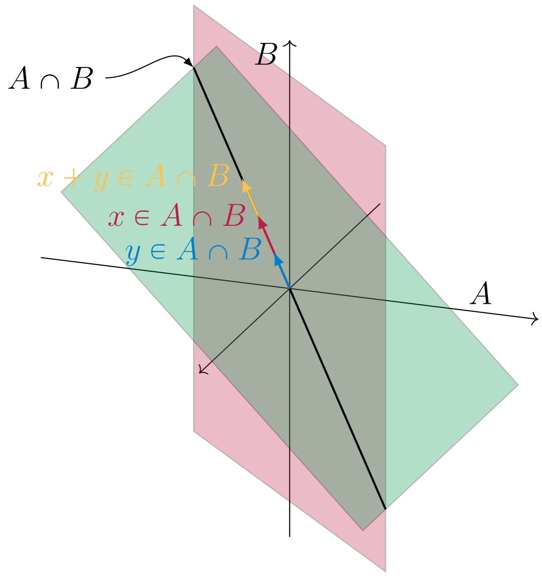
Lemma 3.19. Let V be a vector space and $A, B \subset V$ be two subspaces. Then the *intersection*

$$A \cap B := \{v \in V \mid v \in A \text{ and } v \in B\}$$

is also a subspace of V. More generally, this holds true for any number of subspaces, i.e., if $A_1, A_2 \dots, A_n \subset V$ are subspaces, then so is their joint intersection

$$A_1 \cap \cdots \cap A_n = \{v \in V \mid v \in A_1, v \in A_2, \dots, v \in A_n\}.$$

Proof. We need to make sure that $A \cap B$ satisfies the conditions in Definition 3.17. This is easy enough. For example, the zero vector $0 \in A \cap B$ since $0 \in A$ (since A is a sub*space*) and also $0 \in B$ (since B is also a subspace). Here is a visualization for sums: if $x, y \in A \cap B$, then $x + y \in A \cap B$ since it is both contained in A and also in B.



□

Intersections of subspaces are hugely important to us because of the following example.

Example 3.20. Consider once again a homogenous system as in (3.14). Then, of course, each individual equation of that system is in its own right a homogeneous linear equation, for example

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0$$

By the above, the solution set of that equation is a subspace of $\mathbf{R}^n$, that we denote by $S_1 (\subset \mathbf{R}^n)$. Likewise this is true for all the other individual equations, so we get some subspaces $S_1, \dots, S_m$, one for each equation. The solution set of the *whole* system is then just

$$S_1 \cap S_2 \cap \cdots \cap S_m.$$

(Indeed, a vector $(r_1, \dots, r_n) \in \mathbf{R}^n$ is a solution for the whole system precisely if it is one for the individual equations.)

An important question that we will eventually be able to make more precise and to answer is this:

Question 3.21. Given two subspaces A, B in some vector space V, how "much smaller" can $A \cap B$ be than A and B?

In the above illustration, we will want to articulate the idea that the ambient vector space V is "3-dimensional", A and B are 2-dimensional (i.e., a plane) and $A \cap B$ is 1-dimensional (i.e., a line).

Note that this need not be the case: if $A = B$ is the *same* plane, for example, then certainly $A \cap B = A$ is also 2-dimensional. This relates to the discussion about the intersections of lines in $\mathbf{R}^2$ in Summary 2.12: if $A, B \subset \mathbf{R}^2$ are "1-dimensional" (i.e., lines), their intersection may still be a line, namely if $A = B$. If the ambient vector space V is even larger, for example $V = \mathbf{R}^4$ (which has "dimension 4"), then it is no longer reasonable to write down all possible constellations of how A, B lie in V.

3.4 Further examples of vector spaces

3.4.1 Polynomials

We introduce a number of further examples of vector spaces. Recall that a function $f : \mathbf{R} \to \mathbf{R}$ (i.e., cf. §A, a function that takes as an input a real number x and whose output $f(x)$ is another real number) is called a *polynomial* if it is of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_n, a_{n-1}, \ldots, a_0$ are real numbers. These numbers are called the *coefficients* of f. The *degree* of f is the largest exponent n appearing in f (provided that the coefficient $a_n \neq 0$). Recall from §A that such an expression is abbreviated as

$$f(x) = \sum_{i=0}^{n} a_i x^i.$$

In increasing complexity, a *constant function*

$$f(x) = a$$

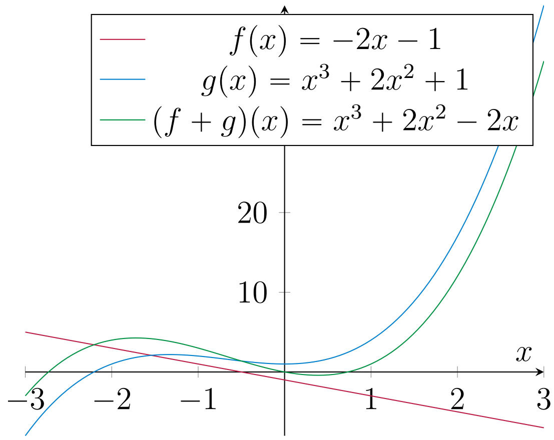
is a polynomial of degree 0 (note $a = a \cdot x^0$);

$$f(x) = a_1 x + a_0$$

is a *linear polynomial* (also known as *linear function*). Its degree is 1 (provided $a_1 \neq 0$). Next,

$$f(x) = a_2 x^2 + a_1 x + a_0$$

is called a *quadratic polynomial* (or *quadratic function*). Its degree is 2 (provided $a_2 \neq 0$; if $a_2 = 0$ then it is a linear polynomial). These types of functions are familiar from high-school.



Definition and Lemma 3.22. The set

$$\mathbf{R}[x] := \{f : \mathbf{R} \to \mathbf{R} \mid f \text{ is a polynomial}\}$$

is a vector space, where we define the sum and scalar multiplication as follows: given two polynomials $f, g \in P$, their *sum* is the function $f + g : \mathbf{R} \to \mathbf{R}$defined by

$$(f + g)(x) := f(x) + g(x),$$

and given a real number $r \in \mathbf{R}$, the scalar multiple is the function $rf : \mathbf{R} \to \mathbf{R}$ defined by

$$(rf)(x) := r \cdot f(x).$$

The set

$$\mathbf{R}[x]^{\leqslant d} := \{\sum_{i=0}^{d} a_i x^i \; | a_0, \dots, a_d \in \mathbf{R}\} (\subset \mathbf{R}[x])$$

of polynomials of *degree* at most d is a subspace of $\mathbf{R}[x]$.

Proof. We have to check the conditions on a vector space (Definition 3.10). As it also happens often in other examples, the most notable condition to check is that the sum and scalar multiple is again an element of the vector space. Here, we need to check that for $f, g \in P$ the function $f + g$ defined above is again a polynomial. Fortunately, this is easy: if $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$

and $g(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0$, then, by definition

$$\begin{aligned}
(f+g)(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 + b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0 \\
&= a_n x^n + b_n x^n + a_{n-1}x^{n-1} + b_{n-1}x^{n-1} + \cdots + a_0 + b_0 \\
&= \underbrace{(a_n + b_n)}_{=:c_n} x^n + \underbrace{(a_{n-1} + b_{n-1})}_{=:c_{n-1}} x^{n-1} + \cdots + \underbrace{(a_0 + b_0)}_{=:c_0} \\
&= c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0.
\end{aligned}$$

Thus, the sum of f and g is another polynomial. Similarly, one verifies that the scalar multiple $r \cdot f$ is a polynomial (check this! what are its coefficients?). With these checks done, one can proceed checking the remaining conditions in Definition 3.10. Checking this is comparatively uninsightful, and will be skipped.

Checking that $\mathbf{R}[x]^{\leqslant d}$ is a subspace amounts to asserting that the 0 polynomial $f(x) = 0$ has degree at most d, and that sums and *scalar* (!) multiples of polynomials of degree $\leqslant d$ have again degree $\leqslant d$. This is clear. □

Remark 3.23. It is also true that the *product* of two polynomials is again a polynomial, but this is not part of what it takes to be a vector space, so we disregard that property at this point.

Remark 3.24. Instead of just polynomials, one can consider more general functions:

$$\begin{aligned}
\mathbf{R}[x] &\subset \{f : \mathbf{R} \to \mathbf{R} \mid f \text{ is differentiable } \} \\
&\subset \{f : \mathbf{R} \to \mathbf{R} \mid \text{ is continuous } \} \\
&\subset \{f : \mathbf{R} \to \mathbf{R} \text{ is any function}\}
\end{aligned}$$

are increasingly large vector spaces, cf. Exercise 3.3. The (huge!) space of all differentiable functions is a key player in analysis.

3.4.2 Direct sums

Definition and Lemma 3.25. Let V, W be two vector spaces. Their *direct sum* is the set

$$V \oplus W := \{(v, w) \mid v \in V, w \in W\}.$$

It is endowed with the addition given by

$$(v, w) + (v', w') := (v + v', w + w')$$

and scalar multiplication given by

$$r \cdot (v, w) := (rv, rw).$$

These operations turn $V \oplus W$ into a vector space.

More generally, the same definition works for finitely many[1] vector spaces $V_1, \dots, V_n$, giving rise to the direct sum $V_1 \oplus \cdots \oplus V_n$.

This is easy to check: revisit the definition of a vector space and see how checking each of the 8 axioms for $V \oplus W$ reduces to using the precise same axioms for V and W. In particular, the zero vector in $V \oplus W$ is the pair $(0_V, 0_W)$, where for clarity 0_V denotes the zero vector in V and 0_W the one in W.

Example 3.26. We have $\mathbf{R}^2 = \mathbf{R} \oplus \mathbf{R}$ and in general

$$\mathbf{R}^n = \underbrace{\mathbf{R} \oplus \cdots \oplus \mathbf{R}}_{n \text{ summands}}.$$

This is clear from the definition of the sum of vectors in $\mathbf{R}^n$ (Definition 3.3) and the scalar multiplication (Definition 3.7).

Note that $V \subset V \oplus W$, by regarding a vector $v \in V$ as the vector $(v, 0_W)$. Likewise we can regard some $w \in W$ as the vector $(0_V, w) \in V \oplus W$. This way, $V \oplus W$ is a vector space that naturally contains both V and W.

Example 3.27. The direct sum $\mathbf{R}^2 \oplus \mathbf{R}$ consists of pairs (v, w) with $v = (x, y) \in \mathbf{R}^2$ and $w \in \mathbf{R}$. Thus, $\mathbf{R}^2 \oplus \mathbf{R} = \{((x, y), w) \mid x, y, w \in \mathbf{R}\}$. We can identify such a pair (consisting of a pair (x, y) and a number w) with a triple (x, y, w). Therefore, $\mathbf{R}^2 \oplus \mathbf{R}$ can be identified with $\mathbf{R}^3$. The sum and scalar multiple on $\mathbf{R}^2 \oplus \mathbf{R}$ as defined in Definition and Lemma 3.25 then reduce to the usual sum and scalar multiple in $\mathbf{R}^3$ as defined in Definition 3.3 and Definition 3.7.

3.4.3 Quotient spaces

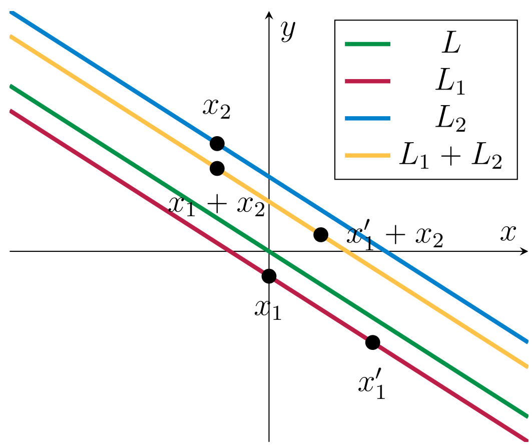
All the examples of vector spaces that we have encountered so far were subspaces of an already given vector space, beginning with some ambient $\mathbf{R}^n$. However, not all vector spaces embed (naturally) in some $\mathbf{R}^n$. To illustrate this, we consider an example of a so-called quotient space. Since a full treatment of this would require a few more basic notions, we only discuss this in a special case:

[1] or even infinitely many

Definition 3.28. Consider $V = \mathbf{R}^2$, the plane, and a line $L \subset W$ through the origin. We define a set

$$V/L := \{ \text{ all lines that are parallel to } L \ \}.$$

(This is read "V modulo L.") For example, L_1 and L_2 are *elements* in that set V/L in the illustration below.



How to define the sum and scalar multiplication on that set V/L? Given $L_1, L_2 \in V/L$, take *any* $x_1 \in L_1$ and any $x_2 \in L_2$. (These are both vectors in $V = \mathbf{R}^2$.) Form the unique line that passes through $x_1 + x_2$ and is parallel to L. Call this line $L_1 + L_2$. Similarly, the scalar multiple $r\dot{L}_1$ is the line passing through $r \cdot x_1$ and parallel to L. What is remarkable is that makes sense, i.e., that the resulting lines do not depent on the choices of x_1, x_2 above. In the illustration below, we indicate two choices for x_1 (the second one being denoted x_1'). The sum $x_1 + x_2$ is clearly different from $x_1' + x_2$, but they do lie on the same line (that is parallel to L). This holds since $x_1' - x_1$ lies in L. Thus

$$(x_1' + x_2) - (x_1 + x_2) = x_1' - x_1$$

also lies in L, and therefore $x_1' + x_2$ and $x_1 + x_2$ lie on the same line that is parallel to L.

With this settled, one can show (withouth much head-ache) that V/L is indeed a vector space. (What is the zero vector in V/L?)

A conceptually important insight is that there is *no natural way* in which this V/L is a subspace of $\mathbf{R}^2$. E.g., one may assign to an element $L_1 \in V/L$, say, the y-coordinate of the intersection of L_1

with the y-axis. But, this idea is ad-hoc and problem-laden (why not take the x-axis instead, and what is worse, what happens if L is in fact the y-axis...).

3.5 Linear combinations

In the sequel, V will always denote a vector space, for example $V = \mathbf{R}^n$.

Definition 3.29. A *linear combination* of vectors $v_1, \dots, v_m \in V$ is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where the $a_1, \dots, a_m$ are arbitrary real numbers.

Example 3.30. If $m = 1$ in the above definition, there is only one vector $v := v_1$. A linear combination of a single vector v is therefore any vector of the form av, with an arbitrary $a \in \mathbf{R}$. In other words it is an arbitrary scalar multiple of that vector.

Example 3.31. More interesting things start happening for two vectors and more. As an example, consider $v_1 = (1, 0, 0)$ and $v_2 = (0, 1, 0)$ in the vector space $\mathbf{R}^3$. Then $(3, 2, 0)$ is a linear combination of these since

$$(3, 2, 0) = 3 \cdot (1, 0, 0) + 2 \cdot (0, 1, 0).$$

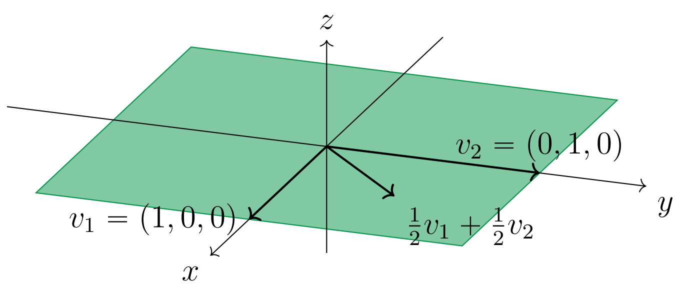
On the other hand, $(0, 0, 1)$ is *not* a linear combination of v_1 and v_2: for arbitrary $a_1, a_2 \in \mathbf{R}$, we compute

$$a_1 v_1 + a_2 v_2 = (a_1, 0, 0) + (0, a_2, 0) = (a_1, a_2, 0).$$

No matter how we choose a_1 and a_2, we always have

$$(a_1, a_2, 0) \neq (0, 0, 1),$$

since the third components of these two vectors are always different. In fact, the linear combinations of v_1 and v_2 are *precisely* the vectors (x, y, z) that satisfy $z = 0$.



Given two vectors $v_1, v_2 \in \mathbf{R}^3$, we will see later (Theorem 3.67) that there will always be some vector $w \in \mathbf{R}^3$ (in fact infinitely many) that is *not* a linear combination of v_1 and v_2. In the above example any vector $w = (x, y, z)$ with $z \neq 0$ has that property. Continuing with that example, the $x-y$-plane inside $\mathbf{R}^3$, i.e., $V := \{(x, y, z) \mid x, y \in \mathbf{R}, z = 0\} = \{(x, y, 0) \mid x, y \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$: it contains $(0, 0, 0)$ and given any two vectors $v, w \in V$ we have $v + w \in V$ and given $r \in \mathbf{R}$, $rv \in V$ (!)(convince yourself this is true!). We can alternatively use Proposition 3.18 to see this is a subspace: V is the solution space of the equation $z = 0$ (in the three variables x, y, z), which is a homogeneous linear equation. The following statement asserts that we always obtain a subspace in this manner.

Lemma 3.32. Let V be a vector space and $v_1, \ldots, v_m \in V$ be any vectors. The set

$$L(v_1, \ldots, v_m) := \{a_1 v_1 + \cdots + a_m v_m \mid a_1, \ldots, a_m \in \mathbf{R}\}$$

of *all* linear combinations of $v_1, \ldots, v_m$ is a subspace of V. It is called the *span* (or sometimes also the *linear hull*) of these vectors.

Proof. We check the three conditions in Definition 3.17. Let us abbreviate $L := L(v_1, \ldots, v_m)$.

(1) The zero vector $0 \in L$ since $0 \cdot v_1 + \ldots 0 \cdot v_m = 0 \cdot (v_1 + \cdots + v_m) = 0$, using properties (4) and (8) in the definitions of a vector space.

(2) Given two vectors $w, u \in L$, we check $w + u \in L$. Since $w \in L$ there are some real numbers $a_1, \ldots, a_m$ such that $w = a_1 v_1 + \cdots + a_m v_m = \sum_{i=1}^m a_i v_i$. Likewise there are real numbers $b_1, \ldots, b_m$

with $u = \sum_{i=1}^m b_i v_i$. This implies

$$\begin{aligned} w + u &= \sum_{i=1}^m a_i v_i + \sum_{i=1}^m b_i v_i \\ &= \sum_{i=1}^m (a_i + b_i) v_i \\ &\in L. \end{aligned}$$

(3) Given $w \in L$ and $r \in \mathbf{R}$, one checks similarly that $rw \in L$ ((!), verify that!). (!)
□

Example 3.33. In Example 3.31, we have

$$L((1,0,0),(0,1,0)) = \{(x,y,0) \mid x,y, \in \mathbf{R}\}.$$

Exercise 3.10 and Exercise 3.11 discuss linear combinations in the vector space $\mathbf{R}[x]^{\leqslant 3}$. The span is closely related to another construction that produces new vector spaces out of given ones:

Definition 3.34. Let V be a vector space and $A, B \subset V$ be two subspaces. The *sum* of A and B is defined as

$$A + B := \{v + w \mid v \in A, w \in B\}.$$

I.e., it consists of *all* possible ways to sum an element in A and an element in B. More generally, given subspaces $A_1, \dots, A_n$ of V, their sum is defined as

$$A_1 + \cdots + A_n := \{v_1 + \cdots + v_n \, | v_1 \in A_1, \dots, v_n \in A_n\}.$$

Lemma 3.35. The sum $A + B$ is then again a subspace of V.

The proof of this is very similar to the one of Lemma 3.32 and will be omitted.

Remark 3.36. Given some vectors $v_1, \dots, v_n \in V$, we have

$$L(v_1, \dots, v_n) = L(v_1) + \cdots + L(v_n).$$

Indeed, both sets are precisely the vectors of the form $a_1 v_1 + \cdots + a_n v_n$ for arbitrary $a_i \in \mathbf{R}$.

Remark 3.37. The sum is *completely different* from the union $A \cup B$ of the two subspaces. We have seen in Example 3.13 that the union is (in general) not even a subspace (just a subset). We have

$$A \cup B \subset A + B,$$

but these two subsets are distinct (unless A or B only consists of the zero vector). To see this inclusion, note that $A \subset A + B$. Indeed, the zero vector $0 \in B$ (since it is a subspace!), and for any $v \in A$, we have $v = v + 0 \in A + B$. Similarly, $B \subset A + B$, and therefore $A \cup B \subset A + B$.

Remark 3.38. The sum $A + B$ is *different* from the direct sum $A \oplus B$. This is already clear from the definition: while the sum makes use of the ambient vector space V, the direct sum $A \oplus B$ is insensitive to A and B both lying in V. Also, it does not "see" to what extent A and B may overlap.

In a spirit similar to Question 3.21 we can ask the following question:

Question 3.39. Given two subspaces $A, B \subset V$ of some larger vector space, how much "bigger" than A and B is the sum $A + B$?

It turns out that Question 3.21 are closely related. Loosely speaking, one can say that $A + B$ "gets bigger" the same way as $A \cap B$ "gets smaller". To give a precise meaning to this one needs the concept of the dimension of a vector space. Understanding the dimension of a vector space requires combining two preliminary notions, that of a generating system and that of linear independence below (Definition 3.46).

Definition 3.40. A collection $v_1, \dots, v_n$ of vectors is a *generating system* if

$$L(v_1, \dots, v_n) = V$$

or, equivalently, if *every* vector $w \in V$ is an appropriate linear combination of these vectors. We also say that these vectors *span* V if this is the case.

Example 3.41. The vectors $e_1 := (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$ up to $e_n = (0, \dots, 0, 1)$ are a generating system. Indeed, each vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ is a linear combination of these, namely

$$x = x_1 \cdot e_1 + \dots + x_n \cdot e_n.$$

Example 3.42. We have observed in Example 3.31 that the vectors $e_1 = (1,0,0)$ and $e_2 = (0,1,0)$ in $\mathbf{R}^3$ are *not* a generating system since they only span the subspace

$$L(e_1, e_2) = \{(x, y, 0) \mid x, y \in \mathbf{R}\}$$

which is not the entire $\mathbf{R}^3$ (e.g., $(0,0,1)$ is missing).

The following example shows that three arbitrary vectors in $\mathbf{R}^3$ need not form a generating set.

Example 3.43. Consider the vectors $v_1 := e_1 = (1,0,0)$, $v_2 = (0,1,1)$ and $v_3 = (2,1,1)$. These three vectors do *not* form a generating set of $\mathbf{R}^3$. In order to show this and to also understand which vectors are precisely in the span $L(v_1, v_2, v_3)$, we consider the following equation, where $w = (x, y, z) \in \mathbf{R}^3$ is a vector and $a_1, a_2, a_3 \in \mathbf{R}$:

$$w = a_1 v_1 + a_2 v_2 + a_3 v_3.$$

Those vectors w that can be written in such a form are in the span, those where no such equation holds are not in the span! This is an equation between two vectors in $\mathbf{R}^3$, i.e., ordered triples. Two such triples are the same precisely if their three components are the same. This leads to the following linear system:

$$\begin{aligned} a_1 \cdot 1 + a_2 \cdot 0 + a_3 \cdot 2 &= x, \\ a_1 \cdot 0 + a_2 \cdot 1 + a_3 \cdot 1 &= y, \\ a_1 \cdot 0 + a_2 \cdot 1 + a_3 \cdot 1 &= z. \end{aligned}$$

In this system a_1, a_2, a_3 are the variables, and x, y, z are parameters (on which the solutions of the system will depend). We form the matrix associated to this linear system, which is

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & x \\ 0 & 1 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right).$$

We apply Gaussian elimination to that matrix, i.e., subtract the second row from the third:

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & x \\ 0 & 1 & 1 & y \\ 0 & 0 & 0 & z - y \end{array}\right).$$

We now distinguish two cases:

- $z - y = 0$ (i.e., $y = z$): in this case the matrix is already in reduced row echelon form (Definition 2.28). The system has solutions, namely the variable a_3 is a free variable, so its value be chosen arbitrarily. Then a_1 and a_2 are uniquely determined by a_3 by the equations

$$\begin{aligned} a_1 + 2a_3 &= x, \\ a_2 + a_3 &= y, \end{aligned}$$

  which gives $a_1 = x - 2a_3$ and $a_2 = y - a_3$. Therefore, for arbitrary $x, y \in \mathbf{R}$, the vectors

$$w = (x, y, y) \in L(v_1, v_2, v_3)$$

  are in the span. They can be expressed as linear combinations

$$w = (x - 2a)v_1 + (y - a)v_2 + av_3,$$

  for an arbitrary $a \in \mathbf{R}$ (this was the a_3 before).

- $z - y \neq 0$ (i.e., $y \neq z$). In this case, we can divide the last equation by $z-y$, which gives the following reduced row-echelon matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & x \\ 0 & 1 & 1 & y \\ 0 & 0 & 0 & 1 \end{array}\right).$$

  According to Method 2.32, the system has *no* solution in this case. Thus, vectors of the form

$$w = (x, y, z) \text{ with } y \neq z$$

  are *not* in the span: $w \notin L(v_1, v_2, v_3)$.

The following method gives a criterion to check whether a given set of vectors generates $\mathbf{R}^n$. We will prove this statement later (Theorem 4.80).

Method 3.44. Let $v_1, \dots, v_m \in \mathbf{R}^n$ be some vectors. Form the matrix

$$A = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

(i.e., each the i-th row of A is precisely the vector v_i, so that $A = (v_{ij})$ if $v_i = (v_{i1}, \dots, v_{in})$.) Bring this matrix into row-echelon form by Gaussian elimination (Method 2.30). Call this resulting matrix B. If B contains n leading ones, then $v_1, \dots, v_m$ span $\mathbf{R}^n$. Otherwise, they don't span $\mathbf{R}^n$.

Corollary 3.45. Fewer than n vectors can *never* span $\mathbf{R}^n$ (since in any event B can at most contain m leading ones).

3.6 Linear independence

Let $v_1, \dots, v_m \in V$ be m vectors in some vector space. Then we have

$$0 \cdot v_1 + \cdots + 0 \cdot v_m = 0 \cdot (v_1 + \cdots + v_m) = 0.$$

This follows from the distributive law and the scalar multiplication of any vector with 0, cf. (4) and (8) in Definition 3.10. So, there is always a "trivial" way to obtain the zero vector from $v_1, \dots, v_m$. We can ask if there are other ways of achieving the zero vector.

Definition 3.46. We say $v_1, \dots, v_m$ are *linearly dependent* if there is a *non-zero* linear combination of these that gives the zero vector. I.e., if there are $a_1, \dots, a_m \in \mathbf{R}$ of *which at least one is non-zero*, such that

$$a_1 v_1 + \cdots + a_m v_m = 0. \tag{3.47}$$

If this is not the case, then we say the vectors are *linearly independent*.

Thus, they are linearly independent if the zero linear combination in (3.6) is the *only* way to obtain the zero vector as a linear combination of $v_1, \dots, v_m$.

Example 3.48. The vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1) \in \mathbf{R}^3$ are linearly independent. To see this, suppose some linear combination equals the zero vector: if

$$a_1 e_1 + a_2 e_2 + a_3 e_3 = (0, 0, 0)$$

then we compute the left hand side as

$$(a_1, 0, 0) + (0, a_2, 0) + (0, 0, a_3) = (a_1, a_2, a_3),$$

so the above equation forces $a_1 = a_2 = a_3 = 0$. This shows that the vectors are linearly independent.

More generally, the same argument shows that

$$e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots 0), \ldots, e_n = (0, \ldots, 0, 1) \in \mathbf{R}^n$$

are linearly independent.

Example 3.49. We revisit the vectors $v_1 := e_1 = (1, 0, 0)$, $v_2 = (0, 1, 1)$ and $v_3 = (2, 1, 1) \in \mathbf{R}^3$ of Example 3.43. These vectors are *not* linearly independent. Indeed, we observe that $v_3 = 2v_1 + v_2$, so that

$$2v_1 + v_2 - v_3 = 0.$$

Example 3.50. The polynomials $1 + x$, $3x + x^2$, $2 + x - x^2$ are linearly independent vectors in $\mathbf{R}[x]^{\leqslant 2}$. To see this, suppose that a linear combination of them equals the zero vector (i.e., the constant polynomial 0):

$$\begin{aligned} 0 &= a_1(1 + x) + a_2(3x + x^2) + a_3(2 + x - x^2) \\ &= a_1 + 2a_3 + (a_1 + 3a_2 + a_3)x + (a_2 - a_3)x^2. \end{aligned}$$

Since this must hold for all $x \in \mathbf{R}$, this forces the following homogeneous linear system:

$$\begin{aligned} 0 &= a_1 + 2a_3 \\ 0 &= a_1 + 3a_2 + a_3 \\ 0 &= a_2 - a_3. \end{aligned}$$

(!) Solving this system (do it (!)) one sees that this only has the trivial solution $a_1 = a_2 = a_3 = 0$. Thus, the polynomials are linearly independent.

The following statement says in some sense that a family of vectors is linearly independent if there is no redundancy among them.

Lemma 3.51. Let $v_1, \ldots, v_m \in V$ be some vectors. They are linearly dependent exactly if (at least) *one* of these vectors can be expressed as a linear combination of the others, i.e., some

$$v_i = a_1 v_1 + \cdots + a_{i-1} v_{i-1} + a_{i+1} v_{i+1} + \cdots + a_m v_m \qquad (3.52)$$

for an appropriate i and appropriate coefficients a_1 etc.

Proof. If (3.52) holds, then

$$a_1v_1 + \cdots + a_{i-1}v_{i-1} + (-1)v_i + a_{i+1}v_{i+1} + \cdots + a_mv_m = 0,$$

so they are linearly dependent.

Conversely, if (3.47) holds, then pick some i such that $a_i \neq 0$ (by assumption this is possible). Then one can subtract a_iv_i and divide by $-a_i$ (which is nonzero, crucially!), giving

$$v_i = \frac{-a_1}{a_i}v_1 + \cdots + \frac{-a_{i-1}}{a_i}v_{i-1} + \frac{-a_{i+1}}{a_i}v_{i+1} + \cdots + \frac{-a_m}{a_i}v_m.$$

This is an equation of the form (3.52). □

The following method decides whether a given set of vectors is linearly independent in $\mathbf{R}^n$. A proof is conveniently done using later results, such as Lemma 4.76.

Method 3.53. Let $v_1, \ldots, v_m \in \mathbf{R}^n$ be some vectors. Form the matrix

$$A = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

(i.e., each the i-th row of A is precisely the vector v_i, so that $A = (v_{ij})$ if $v_i = (v_{i1}, \ldots, v_{in})$.) Bring this matrix into row-echelon form using Gaussian elimination (Method 2.30). Call this resulting matrix B. If B contains m leading ones, then $v_1, \ldots, v_m$ are linearly independent. Otherwise, they are linearly dependent.

Corollary 3.54. More than n vectors can *never* be linearly independent in $\mathbf{R}^n$ (i.e., for $m > n$, any vectors $v_1, \ldots, v_m$ will be linearly dependent, since the matrix B can contain at most n leading ones, being in reduced row-echelon form).

Remark 3.55. This method is very similar to Method 3.44, except that there we asked B to contain n leading ones: this guarantees that $v_1, \ldots, v_m$ span $\mathbf{R}^n$. Having as many leading ones as there are vectors, i.e., m leading ones, instead guarantees that the vectors are linearly independent.

Example 3.56. We revisit the vectors $v_1 := e_1 = (1,0,0)$, $v_2 = (0,1,1)$ and $v_3 = (2,1,1) \in \mathbf{R}^3$ of Example 3.56. The matrix having these vectors as rows is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}.$$

We bring it into reduced row echelon form like so:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

This reduced row-echelon matrix has only 2 leading ones, so the vectors are *not* linearly independent, i.e., they are linearly dependent.

The importance of linearly independent vectors comes from the following result:

Proposition 3.57. Let $v_1, \dots, v_m$ be linearly independent vectors in a vector space V. If some vector v can be expressed as an (ostensibly different) linear combination of those, these presentations must be the same. I.e., if

$$\begin{aligned} v &= a_1 v_1 + \cdots + a_m v_m \text{ and} \\ v &= b_1 v_1 + \cdots + b_m v_m \end{aligned}$$

for appropriate real numbers $a_1, \dots, a_m, b_1, \dots, b_m$, then necessarily we have

$$a_1 = b_1, a_2 = b_2, \dots, a_m = b_m.$$

Proof. Subtracting these two equations from one another (and using the commutativity of addition, and the law of distributivity, cf. Definition 3.10), we obtain

$$\begin{aligned} 0 &= v - v \\ &= (a_1 - b_1) v_1 + \cdots + (a_m - b_m) v_m. \end{aligned}$$

Since the vectors are linearly independent, this implies $a_1 - b_1 = 0$ etc., so that $a_1 = b_1$ etc. □

3.7 Bases

Definition 3.58. A collection of vectors in a vector space

$$v_1, \dots, v_m \in V$$

is called a *basis* if they span V and if they are linearly independent.

Example 3.59. The vectors

$$e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots 0), \dots, e_n = (0, \dots, 0, 1) \in \mathbf{R}^n$$

are a basis, called the *standard basis*. Indeed, we have observed in Example 3.41 and Example 3.48 that they span $\mathbf{R}^n$ and that they are linearly independent.

We try and modify this basis a little bit and see what happens. If we omit one of the vectors and only consider, say

$$e_2 = (0, 1, 0, \dots 0), \dots, e_n = (0, \dots, 0, 1) \in \mathbf{R}^n$$

these do not form a basis: while they are still linearly independent, they do not span $\mathbf{R}^n$.

On the other hand, we now consider

$$e_1, \dots, e_n, v,$$

for an arbitrary vector $v \in \mathbf{R}^n$. These do not form a basis: while they span $\mathbf{R}^n$ (even without the v), they are not linearly independent. Indeed, since $e_1, \dots, e_n$ span $\mathbf{R}^n$, this means that

$$v = a_1 e_1 + \dots + a_n e_n$$

for appropriate $a_1, \dots, a_n \in \mathbf{R}$. According to Lemma 3.51, this means that $e_1, \dots, e_n, v$ are linearly dependent.

Example 3.60. The vectors

$$v_1 = (0, 2, 1), v_2 = (1, 0, 2), v_3 = (-1, 1, 1)$$

form a basis of $\mathbf{R}^3$. To see this, we apply Method 3.53 and Method 3.44:

$$\begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ -1 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{5}{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

This matrix has three leading ones, so the vectors are linearly independent and span $\mathbf{R}^3$, so they form a basis.

Note that this is a *different* basis than e_1, e_2, e_3 considered above.

The following result, which is simply a combination of the definition of generating systems and Proposition 3.57, is often described by saying that a basis gives rise to a *coordinate system* in a vector space.

Proposition 3.61. Let $v_1, \dots, v_m$ be a basis of a vector space V. Then each vector $v \in V$ can be written in a *unique* way as a linear combination

$$v = a_1 v_1 + \cdots + a_m v_m.$$

For some other vector $w = b_1 v_1 + \cdots + b_m v_m$, we have

$$v + w = (a_1 + b_1) v_1 + \cdots + (a_m + b_m) v_m.$$

3.8 The dimension of a vector space

We are all used to referring to the space surrounding us as "3-dimensional", and refer to a plane as "2-dimensional". In this section, which is crucial to linear algebra and, by extension to all applications of linear algebra in physics, engineering and mathematics itself, we make this statement precise.

Theorem 3.62. Let V be a vector space with a basis $v_1, \dots, v_n$. Then any other basis of V also consists of n vectors.

In other words, the number of vectors in a basis does *not* depend on the basis. (Recall from Example 3.60 that the vectors that form a basis may very well be different.)

Definition 3.63. We say that a vector space V has *dimension* n if there is a basis of V with n elements.

Example 3.64. The standard basis of $\mathbf{R}^n$ consists of n elements (Example 3.59), so that

$$\dim \mathbf{R}^n = n.$$

The space of polynomials of degree at most d has a basis $1, x, x^2, \dots, x^d$. These are $d + 1$ polynomials so that

$$\dim \mathbf{R}[x]^{\leqslant d} = d + 1.$$

If V has a basis $v_1, \dots, v_n$ (so that $\dim V = n$) and another vector space W has a basis $w_1, \dots, w_m$ (and $\dim W = m$), then a basis of the direct sum $V \oplus W$ is given by

$$(v_1, 0), \dots, (v_n, 0), (0, w_1), \dots, (0, w_m).$$

These are $n + m$ vectors, so that

$$\dim(V \oplus W) = \dim V + \dim W.$$

Theorem 3.65. Every vector space has a basis.

Remark 3.66. In this course, we only consider vector spaces with a basis consisting of finitely many vectors, as in Definition 3.58. We call such vector spaces *finite-dimensional*.

An example of a vector space not having a finite basis (i.e., an *infinite-dimensional* vector space) is $\mathbf{R}[x]$ (for which a basis is given by the polynomials $1, x, x^2, x^3, \dots$).

The following theorem addresses the question how linearly independent sets can be extended to a basis.

Theorem 3.67. (1) Suppose that some vector space V is spanned by m vectors $v_1, \dots, v_m$ (Definition 3.40). Then a basis of V can be obtained by removing certain vectors among the $v_1, \dots, v_m$. In particular, this says that V *has* a basis and that

$$\dim V \leqslant m$$

(and so, in particular that V is finite-dimensional.)

(2) Every linearly independent set of vectors can be enlarged to a basis by adding appropriate vectors from any given basis of V. (I.e., if $v_1, \dots, v_n$ are linearly independent, and $w_1, \dots, w_m$ is any basis of V, then the $v_1, \dots, v_n$ together with certain vectors among the $w_1, \dots, w_m$ form a basis.) In particular, if $v_1, \dots, v_n$ are linearly independent, then

$$\dim V \geqslant n.$$

(3) If $W \subset V$ is a subspace, then $\dim W \leqslant \dim V$. (In particular, if V is finite-dimensional, then so is W.) Moreover, we have $\dim W = \dim V$ precisely if $W = V$.

(4) For a subspace $W \subset V$, any basis of W can be extended to a basis of V.

Proof. This is proved in any linear algebra textbook, e.g., [Nic95, Theorem 6.4.1] or [Bot21, §1.3]. □

Example 3.68. In $V = \mathbf{R}^3$, consider the four vectors $v_1 = (1, 1, -1)$, $v_2 = (2, 0, 1)$, $v_3 = (-1, 1, -2)$, $v_4 = (1, 2, 1)$. We apply Method 3.44 and Method 3.53 by forming the associated matrix and bringing it into row echelon form:

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ -1 & 1 & -2 \\ 1 & 2 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & -2 & 3 \\ 0 & 2 & -3 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{7}{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(First step: add certain multiples of the first row to the others, second step: multiply second row by $-\frac{1}{2}$ and add multiples to the third and last row, third step: divide the last row by $\frac{7}{2}$.) We can swap the last two rows and obtain a row echelon matrix. This matrix has *three* leading ones, so that the four vectors generate $\mathbf{R}^3$ but are not linearly independent. (We also know $\dim V = 3$, so these four vectors can not be linearly independent by Theorem 3.67(2).) According to Theorem 3.67(1), we can obtain a basis by removing certain vectors among these. Notice that one may not (in general) remove just any arbitrary of the four vectors. In this example,

- the first three vectors v_1, v_2, v_3 do *not* form a basis,
- however v_1, v_2, v_4 *do* form a basis.

Indeed, this holds since in the above matrix, we remove either the last row, which brings us to

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

This tells us that these three vectors are (still) not linearly independent (and don't span $\mathbf{R}^3$). By contrast, removing the third row, gives

$$\begin{pmatrix} v_1 \\ v_2 \\ v_4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

which has three leading ones, so these three vectors form a basis of $\mathbf{R}^3$.

Corollary 3.69. Let V be a vector space with $\dim V = n$. Let n vectors be given: $v_1, \dots, v_n$. These vectors are linearly independent if and only if they span V.

Proof. This follows from the theorem above. For example, suppose they span V. If they are not linearly independent, then some v_i lies in the span of the remaining vectors. Thus V is the span of all vectors but v_i so that $n - 1 \geqslant \dim V$ by Theorem 3.67(1). This is a contradiction to our assumption.

The converse implication is proved similarly. □

Remark 3.70. If $V = \mathbf{R}^n$, then Corollary 3.69 aligns well with Method 3.44 vs. Method 3.53: we consider the matrix

$$A = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} \rightsquigarrow B$$

and bring it into row echelon form, denoted B. Note that (A and) B are $n \times n$-matrices. Thus, the vectors $v_1, \dots, v_n$ span $\mathbf{R}^n$ if and only if B has n leading ones, which happens if and only if $v_1, \dots, v_n$ are linearly independent.

Example 3.71. Let $a \in \mathbf{R}$ be a fixed real number. Consider the vector space $\mathbf{R}[x]^{\leqslant d}$. The polynomials

$$v_0(x) = (x-a)^0 = 1, v_1(x) = (x-a), \dots, v_d(x) = (x-a)^d$$

are linearly independent. To see this, suppose

$$0 = a_0 v_0 + a_1 v_1 + \cdots + a_d v_d.$$

Note that v_d has degree d, all the remaining ones have degree $\leqslant d-1$. Thus, looking at the coefficient for x^d, we see $a_d = 0$. Continuing this, we note that

$$0 = a_0 v_0 + a_1 v_1 + \cdots + a_{d-1} v_{d-1}$$

forces $a_{d-1} = 0$ (by looking at the coefficient of x^{d-1}). Repeating this argument, one sees that $a_0 = \cdots = a_d = 0$.

We know $\dim \mathbf{R}[x]^{\leqslant d} = d + 1$ (Example 3.64). Thus, by Corollary 3.69, these polynomials $v_0, \ldots, v_d$ form a basis. According to Proposition 3.61, *any* polynomial $f(x)$ of degree $\leqslant d$ therefore can be *uniquely* written as

$$f(x) = a_0 + a_1(x - a) + \cdots + a_d(x - a)^d.$$

Thus, every polynomial can be expressed as a sum of powers of $x - a$. (By definition of a polynomial, it can certainly be expressed as a sum of powers of $x - 0 = x$.) The precise values of a_d are closely related to the *Taylor series* familiar from analysis.

3.8.1 Dimensions of sums and intersections

In this section, we give an answer to Question 3.21 and Question 3.39. Colloquially, the possible failure of $A + B$ being "as large as possible" (i.e., having the maximum possible dimension, namely $\dim A + \dim B$) is closely related to the possible failure of $A \cap B$ being "as small as possible." Before stating that, we note another consequence of Theorem 3.67.

Corollary 3.72. Suppose $A, B \subset V$ are two subspaces with $\dim A = m$ and $\dim B = n$. Then

$$\dim(A + B) \leqslant \dim A + \dim B.$$

(Here, at the left + denotes the sum of the two subspaces (Definition 3.34), while at the right it is the sum of the two dimensions.)

Proof. If $v_1, \ldots, v_m$ is a basis of A and $w_1, \ldots, w_n$ is a basis of B, then they in particular span A, resp. B. Thus, $A + B$ is spanned by

$$v_1, \ldots, v_m, w_1, \ldots, w_n.$$

These are $m + n$ vectors. According to Theorem 3.67(1), this implies

$$\dim(A + B) \leqslant m + n.$$

Theorem 3.73. Suppose $A, B \subset V$ are two subspaces of a vector space. Then

$$\dim(A \cap B) + \dim(A + B) = \dim A + \dim B.$$

This is a special case of a more general theorem, the so-called *rank-nullity theorem* (Theorem 4.26). We illustrate it at the hand of subspaces in $V = \mathbf{R}^2$. If $A \subset V$ is a subspace, then exactly one of the following three cases occurs:

- $\dim A = 0$. This means that A just consists of the zero vector: $A = \{0\}$.
- $\dim A = 1$. This means that there is a basis of A consisting of a single vector $v \in A$. Since v is linearly independent, we have $v \neq 0$ (otherwise $1 \cdot v = 0$ is a non-trivial linear combination giving the zero vector). Since v spans A, this means $A = \{av \mid a \in \mathbf{R}\}$. Thus, A is the line spanned by the (non-zero) vector v.
- $\dim A = 2$. In this case we necessarily have $A = \mathbf{R}^2$ by Theorem 3.67(3).

Of course, for another subspace B the same three cases apply. If $A = \{0\}$, then $A \cap B = \{0\}$ and $A + B = B$, so in this case the dimension formula (4.35) just reads

$$\dim(\{0\}) + \dim B = \dim(\{0\}) + \dim B,$$

which does not give anything interesting. Similarly, if $A = \mathbf{R}^2$, then $A \cap B = B$ and $A + B = \mathbf{R}^2$, so the dimension formula reads

$$\dim B + \dim \mathbf{R}^2 = \dim \mathbf{R}^2 + \dim B.$$

Again, this is tautological. The interesting case is therefore when $\dim A = 1$ and, by symmetry, $\dim B = 1$. Thus both A and B are lines, passing through the origin, in $\mathbf{R}^2$. We distinguish two cases:

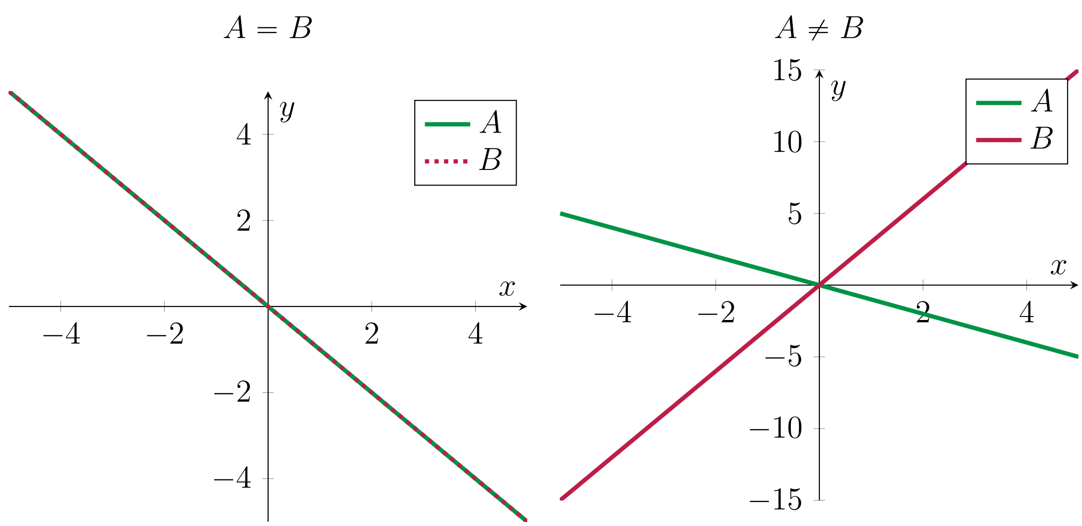
- $A = B$. In this case $A \cap B = A$, $A + B = A$, so the formula reads

  $$1 + 1 = 1 + 1,$$

  which is true.

- $A \neq B$. In this case $A \cap B = \{0\}$, since the lines are distinct and therefore only interesect at the origin. Then the formula says
$$0 + \dim(A + B) = 1 + 1 = 2.$$
Thus $\dim(A + B) = 2$, which means that $A + B = \mathbf{R}^2$, again using Theorem 3.67(3).

Here is a picture of the two cases:



Definition 3.74. Let $A, B \subset V$ be two subspaces. We say "the sum $A + B$ is a direct sum" if $\dim A + \dim B = \dim A + B$.

In other words, $\dim(A + B)$ needs to be as large as possible. In the example of two lines, i.e., $\dim A = \dim B = 1$, the sum is direct precisely if $A + B = \mathbf{R}^2$.

Example 3.75. In $V = \mathbf{R}^3$, consider subspaces $A, B \subset \mathbf{R}^3$ with $\dim A = 1$ and $\dim B = 2$. Thus, geometrically, A is a line passing through the origin and B is a plane passing through the origin. We have
$$0 \subset A \cap B \subset A.$$
This means that
$$0 \leqslant \dim(A \cap B) \leqslant \dim A = 1.$$
We distinguish two cases:

- $A \subset B$. Equivalently, $A \cap B = A$ or, yet equivalently,

$$\dim(A \cap B) = 1.$$

- $A \not\subset B$. In this case $A \cap B \subsetneq A$. Since $A \cap B$ is a subspace of strictly smaller dimension, this implies $A \cap B = \{0\}$. Thus,

$$\dim(A \cap B) = 0.$$

To summarize, a line A and a plane B (both passing through the origin) in $\mathbf{R}^3$ intersect either in a point or in a line.

Example 3.76. Consider $V = \mathbf{R}^3$ and two subspaces $A, B \subset \mathbf{R}^3$ of dimension 2. Then the formula reads

$$\dim(A \cap B) = 2 + 2 - \dim(A + B).$$

We have in any event $A, B \subset A + B \subset \mathbf{R}^3$, which implies

$$2 \leqslant \dim(A + B) \leqslant 3.$$

We consider two cases:

- $A = B$. In this case $A \cap B = A$ and $A + B = A$, which both have dimension 2.

- $A \neq B$. In this case $A \cap B \subsetneq A$, so $A \cap B$ has dimension < 2. This means that $\dim(A + B) = 3$, and therefore

$$\dim(A \cap B) = 1.$$

We summarize this as follows: two planes A, B passing through the origin in $\mathbf{R}^3$ intersect either in a plane (this happens precisely if $A = B$), or they interesect in a line (this happens precisely if $A \neq B$).

If the ambient vector space has dimension $\geqslant 4$, and $\dim A, \dim B \geqslant 2$, then the possible dimensions of $\dim(A \cap B)$ and $\dim(A + B)$ are more varied, so we refrain from making a similar list.

3.9 Exercises

Exercise 3.1. Let $V = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}$. (Thus, $V = \mathbf{R}^3$.) We use the regular addition of vectors. However, in contrast to the regular scalar multiplication (Definition 3.7), we now use the following. Decide in each case whether this turns V into a vector space:

- $r \cdot (x, y, z) = (rx, y, rz)$,
- $r \cdot (x, y, z) = (0, 0, 0)$,
- $r \cdot (x, y, z) = (2rx, 2ry, 2rz)$.

Exercise 3.2. Let $V \subset \mathbf{R}^2$ be a subspace. Which of the following statements are correct?
(1) V contains at least one element.

(2) V contains at least two elements.

(3) V contains the zero vector $(0, 0)$.

(4) If $v, w \in V$ then also $v - w \in V$.

Exercise 3.3. Using basic properties of differentiable functions from your calculus class, show that the space

$$\{f : \mathbf{R} \to \mathbf{R} \mid f \text{ is differentiable }\}$$

is a vector space (with the sum and scalar multiple defined as in (3.22) and (3.22)).

Hint: structure your thinking as in Definition and Lemma 3.22.

Exercise 3.4. Give an example of two subspaces $V, W \subset \mathbf{R}^2$ such that their *union*

$$V \cup W = \{x = (x_1, x_2) \in \mathbf{R}^2 \mid x \in V \text{ or } x \in W\}$$

is *not* a subspace.

Hint: Example 3.13.

Also give an example of two subspaces $V, W \subset \mathbf{R}^2$, where the union $V \cup W$ is a subspace.

Hint: be very lazy and minimalistic. What is the smallest subspace you can come up with?

Exercise 3.5. Determine in each case whether $w \in \mathbf{R}^4$ lies in the span of v_1 and v_2. If so, name at least one linear combination of v_1 and v_2 that equals w; otherwise explain why there is no such linear combination.
(1) $w = (2, -1, 0, 1)$, $v_1 = (1, 0, 0, 1)$, $v_2 = (0, 1, 0, 1)$
(2) $w = (1, 2, 15, 11)$, $v_1 = (2, -1, 0, 2)$, $v_2 = (1, -1, -3, 1)$
(3) $w = (2, 5, 8, 3)$, $v_1 = (2, -1, 0, 5)$, $v_2 = (-1, 2, 2, -3)$

Exercise 3.6. Determine whether the following vectors span $\mathbf{R}^4$:
(1) $(1, 1, 1, 1)$, $(0, 1, 1, 1)$, $(0, 0, 1, 1)$, $(0, 0, 0, 1)$
(2) $(1, 3, -5, 0)$, $(-2, 1, 0, 0)$, $(0, 2, 1, -1)$, $(1, -4, 5, 0)$

Exercise 3.7. Determine whether the following vectors are linearly independent:
(1) $v_1 = (1, -1, 0)$, $v_2 = (3, 2, -1)$, $v_3 = (3, 5, -2)$ in $V = \mathbf{R}^3$,
(2) $v_1 = (1, 1, 1)$, $v_2 = (1, -1, 1)$, $v_3 = (0, 0, 1)$ in $V = \mathbf{R}^3$,
(3) $(1, -, 1, 1, -1)$, $(2, 0, 1, 0)$, $0, -2, 1, -2)$ in $\mathbf{R}^4$,
(4) $(1, 1, 0, 0)$, $(1, 0, 1, 0)$, $(0, 0, 1, 1)$ and $(0, 1, 0, 1)$ in $\mathbf{R}^4$.

Exercise 3.8. Name three vectors $v_1, v_2, v_3 \in \mathbf{R}^2$ such that:

- v_1, v_2 are linearly independent,
- v_1, v_3 are linearly independent, and
- v_2, v_3 are linearly independent, but
- v_1, v_2, v_3 are *not* linearly independent.

Exercise 3.9. Consider the vector space $V = \mathbf{R}[x]^{\leqslant 3}$ of polynomials of degree at most 3. Decide which of the following subsets of V is a subspace:
(1) $\{f \mid f \in V, f(2) = 1\}$,
(2) $\{x \cdot f \mid f \in \mathbf{R}[x]^{\leqslant 2}\}$,
(3) $\{x \cdot f + (1 - x)g \mid f, g \in \mathbf{R}[x]^{\leqslant 2}\}$,
(4) $\{f \mid f \in \mathbf{R}[x]^{\leqslant 3}, f(0) = 0\}$.

Exercise 3.10. Express the following polynomials as linear combinations of $x + 1$, $x - 1$ and $x^2 - 1$ (in $\mathbf{R}[x]^{\leqslant 2}$): $x^2 + 4x - 2$, x, $40 - x^2$.

Exercise 3.11. Is the following sentence correct? "In $\mathbf{R}[x]^{\leqslant 3}$, the polynomial $f(x) = \frac{1}{4}x^3 + 3x + 1$ is a linear combination of the polynomials x^2, x and 1 since $f(x) = \frac{x}{4} \cdot x^2 + 3 \cdot x + 1$."

Exercise 3.12. Express each of the three standard basis vectors e_1, e_2, e_3 as a linear combination of the basis vectors in Example 3.60.

Exercise 3.13. (Solution at p. 235) Consider $A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 2 \\ 3 & 5 \end{pmatrix}$ in the vector space of 2×2-matrices. Is $C = \begin{pmatrix} -1 & 0 \\ 2 & 4 \end{pmatrix}$ a linear combination of A and B?

Exercise 3.14. In the vector space $\mathrm{Mat}_{2\times 3}$ of 2×3-matrices, we consider the set

$$T = \left\{ \begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \end{pmatrix} \mid x_1 + x_4 + x_6 = 0, x_1 + x_4 + x_3 + x_5 = 0 \right\}.$$

(1) Decide whether T is a subspace of $\mathrm{Mat}_{2\times 3}$.

(2) Find all the vectors (i.e., matrices) in T.

(3) Find some vectors such that $T = L(v_1, v_2, v_3, v_4)$.

Exercise 3.15. (Solution at p. 236) In $\mathbf{R}^4$ consider the subset

$$S = \{(x, y, z, t) \mid x + y + z + t = 0\}.$$

(1) Decide whether S is a subspace of $\mathbf{R}^4$.

(2) Find all the vectors (i.e., matrices) in S.

(3) Find some vectors such that $S = L(v_1, v_2, v_3)$.

Exercise 3.16. (Solution at p. 236) Consider the following two subspaces of $\mathbf{R}^4$:

$$S = L((1, -1, 0, 1), (2, 1, -2, 0), (0, 0, 1, 1))$$

and T, which is the solution set of the system

$$\begin{aligned} 2x_1 - x_2 - 3x_4 &= 0 \\ 2x_1 + x_3 + x_4 &= 0. \end{aligned}$$

Determine $S \cap T$.

Exercise 3.17. Consider the following two subspaces of $\mathbf{R}^4$:

$$W = L((1,0,1,0),(2,0,0,0),(0,-3,-1,-1))$$

and T given by the solution set of the system

$$\begin{aligned} x_1 - x_2 &= 0 \\ x_1 + x_2 + x_3 &= 0. \end{aligned}$$

Determine $T \cap W$.

Exercise 3.18. Show that

- $\mathbf{R}^2 = L((1,1),(2,-1))$,
- $\mathbf{R}^2 = L((0,-2),(1,1))$.

Exercise 3.19. Is $(1,5,0) \in \mathbf{R}^3$ a linear combination of $v_1 = (1,1,0)$, $v_2 = (2,0,1)$ and $v_3 = (0,3,-1)$?

(I.e., are there $a_1, a_2, a_3 \in \mathbf{R}$ such that $\alpha_1 v_1 + a_2 v_2 + a_3 v_3 = (1,5,0)$?)

Exercise 3.20. Express the following polynomials as $f(x) = \sum_{i=0}^4 a_i (x-1)^4$:
(1) $f(x) = x^4$,
(2) $f(x) = x^3$,
(3) $f(x) = x^3 - 3x^2 + 4x + 2$.

Exercise 3.21. Let $a, b \in \mathbf{R}$ be two *distinct* numbers. Show that the polynomials $x - a$ and $x - b$ are a basis of $\mathbf{R}[x]^{\leqslant 1}$.

Exercise 3.22. In $\mathbf{R}^4 = \{(x,y,z,t) \mid x,y,z,t \in \mathbf{R}\}$ consider the subspace $W_1 \subset \mathbf{R}^4$ given by the solutions of the system

$$\begin{aligned} y + t &= 0, \\ y + z &= 0. \end{aligned}$$

Also consider the subspace $W_2 = L((0,1,-1,0))$.

Determine a basis and the dimension of W_1. Describe $W_1 \cap W_2$.

Exercise 3.23. Let $k \in \mathbf{R}$ be an arbitrary real number. Consider the subspace

$$W_k := L((1,0,-1,0),(1,1,0,1),(1,2,k,1)) \subset \mathbf{R}^4.$$

(1) For all $k \in \mathbf{R}$, find a basis of W_k and determine $\dim W_k$.

(2) For which $k \in \mathbf{R}$ is $(-1, 1, 1, 1) \in W_k$?

Exercise 3.24. Recall that the dimension of the space $\mathrm{Mat}_{2\times 3}$ of 2×3-matrices is 6.

(1) Consider $W = \left\{ \begin{pmatrix} a & a+b & b \\ 0 & 0 & b \end{pmatrix} \mid a, b \in \mathbf{R} \right\}$. Confirm that W is a subspace of $\mathrm{Mat}_{2\times 3}$. Determine $\dim W$.

(2) Let $V := \left\{ \begin{pmatrix} c & 0 & -c \\ 0 & 0 & -c \end{pmatrix} \mid c \in \mathbf{R} \right\}$. Determine (i.e., determine a basis and the dimension of) $V \cap W$.

Exercise 3.25. (Solution at p. 237) Consider the following subspaces of $\mathbf{R}^3$:

$$W_1 := L((1,0,1),(2,1,0))$$
$$W_2 := L((-1,1,1),(0,3,0)).$$

(1) Determine (i.e., determine a basis and the dimension of) $W_1 \cap W_2$.

(2) Determine $W_1 + W_2$.

Exercise 3.26. Consider the following subspaces of $\mathbf{R}^4$:

$$W_1 := L((1,1,1,2),(2,0,3,5))$$
$$W_2 := L((1,1,0,1),(0,2,-2,-2)).$$

As in the previous exercise, determine $W_1 \cap W_2$ and $W_1 + W_2$.

Exercise 3.27. (Solution at p. 239) Consider the subspace

$$W = L(\underbrace{(1,0,1,0)}_{=v_1}, \underbrace{(2,0,1,1)}_{=v_2}, \underbrace{(0,0,1,3)}_{=v_3}).$$

(1) Find a basis of W and determine $\dim W$.

(2) Find a vector $v \in \mathbf{R}^4$ such that

$$W \subsetneq L(v_1, v_2, v_3, v).$$

What is $\dim L(v_1, v_2, v_3, v)$?

Exercise 3.28. (Solution at p. 240) Consider the vectors in $\mathbf{R}^4$, where $t \in \mathbf{R}$:

$$\begin{aligned} u_1 &= (1, 0, -1, 2) \\ u_2 &= (1, 0, 0, 1) \\ u_3 &= (2, 0, -1, 3) \\ u_4 &= (4, t, -2, 6). \end{aligned}$$

(1) Let $U_t = L(u_1, u_2, u_3, u_4)$ be the subspace spanned by these vectors (where the last vector depends on $t \in \mathbf{R}$). Find the values of t such that
$$\dim U_t = 2.$$

(2) Consider $t = 1$ from now on. Verify $\dim U_1 = 3$ and find a basis of U_1.

(3) Let $W \subset \mathbf{R}^4$ be the subspace given by the equations
$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 - 3x_4 &= 0. \end{aligned}$$
Determine $\dim W$ and $\dim U_1 \cap W$.

Exercise 3.29. Consider the subspace $U_t \subset \mathbf{R}^4$ spanned by the four vectors

$$\begin{aligned} v_1 &= (1, 0, 0, 1) \\ v_2 &= (-1, 1, 2, 3) \\ v_3 &= (0, 1, 2, 4) \\ v_4 &= (t, 2, 4, 8). \end{aligned}$$

Here, $t \in \mathbf{R}$ is an arbitrary real number.

(1) Find the values of t, such that $\dim U_t = 2$.

(2) Consider from now on $t = 1$. Determine $\dim U_1$.

(3) Let $W \subset \mathbf{R}^4$ be the subspace given by the equations
$$\begin{aligned} x_1 - x_2 &= 0 \\ x_2 - x_3 &= 0. \end{aligned}$$
Determine a basis and the dimension of W and of $W \cap U_1$.

Exercise 3.30. (Solution at p. 243) Let $f\colon \mathbf{R}^3 \to \mathbf{R}^4$ be the following linear map:

$$f(x, y, z) = (-y + 2z,\ 2x + y,\ -x + 2y - 5z,\ x + 2y - 3z)$$

(1) Compute a basis of $\ker f$ and a basis of $\operatorname{im} f$.

(2) Let $W \subset \mathbf{R}^4$ be the subspace defined by the equation $x_2 - 3x_3 = 0$. Compute the dimension and a basis of W.

(3) We put $U = \Im f$. Compute a basis of $U \cap W$ and a basis of $U + W$.

(4) Determine for which values of a and b there exists a vector $v \in \mathbf{R}^3$ such that $f(v) = (a, 4, 3, b)$.

Chapter 4

Linear maps

Mathematical objects gain a lot of richness when they can be related to each other. In linear algebra, the objects of interest are vector spaces, and the way the relate to each other is by means of linear maps. The word "map" is being used as a synonym to the word "function".

4.1 Definition and first examples

Definition 4.1. Let V, W be two vector spaces. A function $f : V \to W$ is called *linear* (or a *linear map*, or a *linear transformation*) if it satisfies the following conditions:

$$f(v + v') = f(v) + f(v') \qquad \text{for all } v, v' \in V \text{ and} \qquad (4.2)$$

$$f(av) = af(v) \qquad \text{for all } a \in \mathbf{R}, v \in V. \qquad (4.3)$$

The vector space V is called the *domain* of f, W is called the *codomain* of f.

Remark 4.4. These two conditions can be squeezed into one condition, by requiring that

$$f(av' + a'v') = af(v) + a'f(v'),$$

for all $a, a' \in \mathbf{R}$ and all $v, v' \in V$. This can be paraphrased by saying that f preserves linear combinations.

Using that $0 \cdot v = 0_V$ (the zero vector in V), the above condition implies that

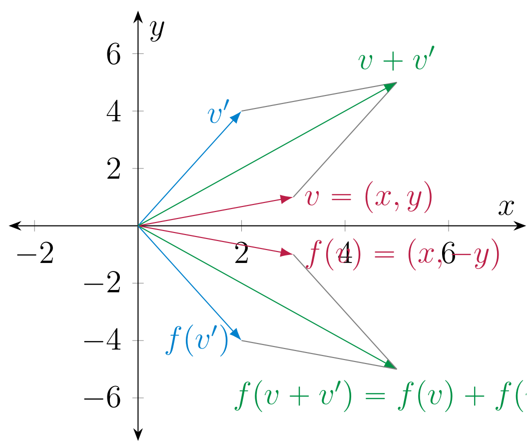
$$f(0_V) = f(0 \cdot v) = 0 \cdot f(v) = 0_W.$$

Thus, for a linear map, the zero vector of V is mapped to the zero vector in W.

Example 4.5. The map $f : \mathbf{R}^2 \to \mathbf{R}^2$, $f(x,y) := (x,-y)$ (i.e., *reflection* at the x-axis) is linear. This can be proven very simply algebraically: for (4.2): if $v = (x,y)$ and $v' = (x',y') \in \mathbf{R}^2$, then

$$f(v+v') = f((x+x', y+y')) = (x+x', -y-y') = (x,-y)+(x',-y') = f(v)+f(v').\blacksquare$$

Checking (4.3) is similarly simple. The linearity of the map can also be visualized geometrically:



We will soon regard the preceding example as a special case of the multiplication of a vector with a matrix, namely in this case the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, cf. §4.2.

Example 4.6. The map

$$D : \mathbf{R}[x] \to \mathbf{R}[x], D(f) := f',$$

i.e., the *derivative* of f, is linear. This is true because we have the formulae (proven in calculus)

$$(f+g)'(x) = f'(x) + g'(x), (af)'(x) = af'(x).$$

Alternatively, one may use that the derivative of a polynomial $f(x) = \sum_{n=0}^{d} a_n x^n$ is given by $f'(x) = \sum_{n=1}^{d} n a_n x^{n-1}$. Then, for

$g = \sum_{n=0}^d b_n x^n$, we check (4.2), say:

$$\begin{aligned}(f+g)'(x) &= \left(\sum_{n=0}^d (a_n+b_n)x^n\right)' \\ &= \sum_{n=1}^d n(a_n+b_n)x^{n-1} \\ &= \sum_{n=1}^d n a_n x^{n-1} + \sum_{n=1}^d n b_n x^{n-1} \\ &= f'(x) + g'(x).\end{aligned}$$

Here are a few slightly more abstract examples of linear maps, in which V is an arbitrary vector space.

Example 4.7.

- The *identity map* $\mathrm{id} := \mathrm{id}_V : V \to V$ which is given by $\mathrm{id}(v) := v$ is linear.
- For some other vector spaces W, the *zero map* $0 : V \to W$ is the map sending every vector v to 0_W. It is linear.
- For any real number $a \in \mathbf{R}$, the map given by scalar multiplication $V \to V$, $v \mapsto a \cdot v$ is linear. This follows from the conditions (4) and (6) in the definition of a vector space (Definition 3.10).

Non-Example 4.8.

- The map $f : \mathbf{R} \to \mathbf{R}$, $f(x) := x^2$ is *not* linear. Indeed, $f(x+y) = (x+y)^2 = x^2 + 2xy + y^2 \neq x^2 + y^2 = f(x) + f(y)$. Also $f(ax) = a^2x^2 \neq ax^2 = af(x)$.
- The map $f : \mathbf{R} \to \mathbf{R}$, $f(x) := x + 1$ is *not* linear since again

  $$f(x+y) = x + y + 1 \neq (x+1) + (y+1) = f(x) + f(y).$$

  Thus, (4.2) is violated. Also (4.3) is violated: $f(ax) = ax + 1 \neq a(x+1) = af(x)$.

4.2 Multiplication of a matrix with a vector

In this section, we define the multiplication of a matrix with a vector and show how this gives rise to a linear map. This is an extremely important way to construct linear maps.

Definition 4.9. Let $A = (a_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ (cf. Notation 2.22) be an $m \times n$-matrix and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ be a $n \times 1$-matrix, i.e., a row vector with n columns. The product of A with v is the $m \times 1$-vector

$$Av := \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{pmatrix}.$$

Thus, the i-th entry of the (column) vector Av is computed by traversing the i-th row of A and multiplying each entry of that row with the corresponding entry of v.

Example 4.10. Here are two concrete examples:

$$\begin{pmatrix} 4 & -1 \\ 2 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \cdot 3 - 1 \cdot 4 \\ 2 \cdot 3 + 1 \cdot 4 \\ 0 \cdot 3 - 2 \cdot 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 10 \\ -8 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 3 \cdot 2 + (-2) \cdot (-1) \\ \\ \\ \end{pmatrix} = \begin{pmatrix} 9 \\ \\ \\ \end{pmatrix}.$$

It makes perfectly good sense to consider matrices whose entries are variables. Compute:

$$\begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \cdot x + 3 \cdot y + (-2) \cdot y \\ \\ \\ \end{pmatrix} = \begin{pmatrix} x + 3y - 2z \\ \\ \\ \end{pmatrix}.$$

Thus, the equation (of column vectors consisting of 3 rows)

$$\begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ -2 \end{pmatrix}$$

is a *very convenient* way to write down the linear system

$$\begin{aligned} x + 3y - 2z &= 1 \\ y &= 4 \\ x - z &= -2. \end{aligned}$$

This shows that the product of matrices with column vectors is very useful in enconding linear systems. We record this observation in the due generality:

Observation 4.11. Let

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

be an $m \times n$-matrix and

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

be a column vector with n rows and

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

be a column vector with m rows. Then the equation

$$Ax = b$$

is equivalent to the linear system (in the unknowns $x_1, \dots, x_n$, consisting of m equations)

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n &= b_m. \end{aligned}$$

4.2.1 The case of 2×2-matrices

The process of multiplying a matrix with a column vector is also geometrically very important. We now investigate this in more detail in the case where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathrm{Mat}_{2\times 2}.$$

For a column vector $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ the product is, according to Definition 4.9,

$$Av = \begin{pmatrix} a_{11}v_1 + a_{12}x_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix}. \tag{4.12}$$

In keeping with traditional notation from geometry, we will instead write the vector v as $\begin{pmatrix} x \\ y \end{pmatrix}$, in which case

$$Av = \begin{pmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{pmatrix}.$$

It is useful to organize this situation into a function, namely the function that sends the vector v to the vector Av. We obtain a function

$$f : \mathbf{R}^2 \to \mathbf{R}^2, v \mapsto Av \text{ (read “}v\text{ maps to }Av\text{”.)}$$

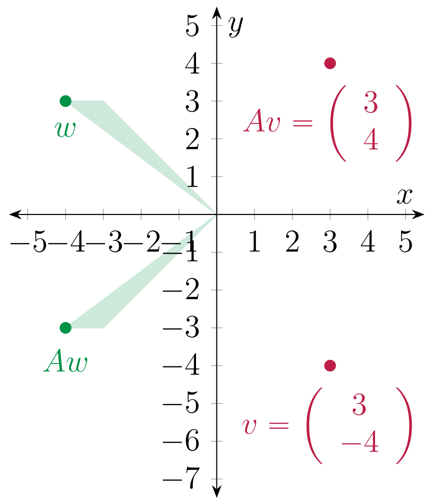
Of course, since Av depends on the entries of A, so does this function f.

Reflections

Example 4.13. We consider $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. According to the above we have

$$Av = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

We plot a few points v and the corresponding Av:



Thus, geometrically, Av is the point v reflected along the x-axis.

Rescalings

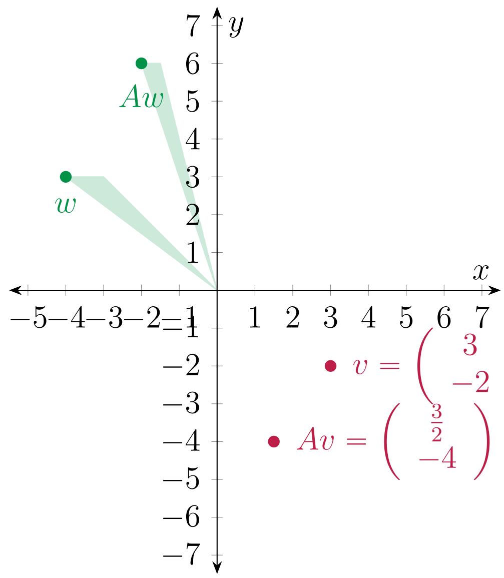
Example 4.14. The matrix $A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$ describes the map that compresses everything in the x-direction by the factor $\frac{1}{2}$, and leaves the y-direction untouched.

Example 4.15. If r, s are two real numbers,

$$A = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}$$

rescales the x-direction by a factor r (so it shrinks for $r < 1$ and enlarges for $r > 1$) and rescales the y-direction by a factor s.

For $A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix}$, this looks as follows:

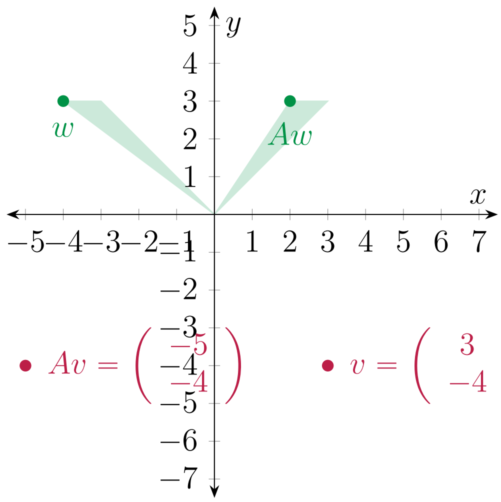


Shearing

Example 4.16. For a fixed real number r, the matrix

$$A = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$$

sends v to $Av = \begin{pmatrix} x + ry \\ y \end{pmatrix}$. Thus it is a shearing operation. In the following picture $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$.



Rotations

We now consider rotations.

Example 4.17. For $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, the vector $Av = \begin{pmatrix} -y \\ x \end{pmatrix}$. Geometrically, the function $v \mapsto Av$ is a counterclockwise rotation by 90°.

For $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, the vector $Av = \begin{pmatrix} -x \\ y \end{pmatrix}$ so the function $v \mapsto Av$ describes a counterclockwise rotation by 180° (or, what is the same, a clockwise rotation by 180°).

For more general rotations, we use basic properties of the trigonometric functions, e.g., as recalled in §B.

Example 4.18. In general, for any $r \in \mathbf{R}$ the matrix

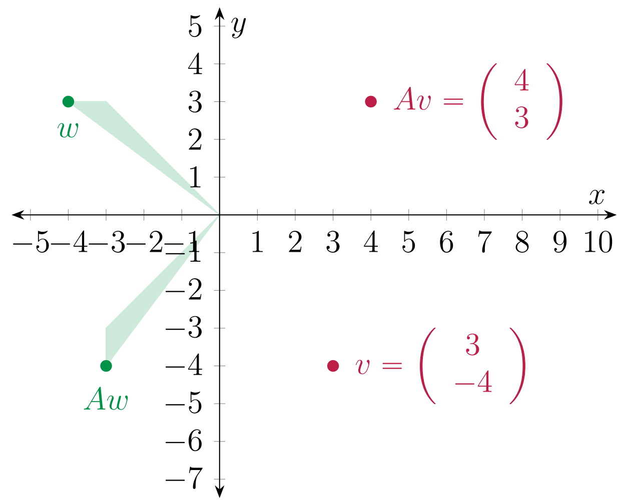
$$A = \begin{pmatrix} \cos r & -\sin r \\ \sin r & \cos r \end{pmatrix}$$

is such that the function

$$v \mapsto Av = \begin{pmatrix} \cos rx - \sin ry \\ \sin rx + \cos ry \end{pmatrix}$$

is a (counter-clockwise) rotation by r. For this reason, A is called a *rotation matrix*.

In the following illustration, $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.



We regard a vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ as an element of $\mathbf{R}^n$. (Thus, instead of using the notation $(v_1, \dots, v_n)$ for an ordered tuple, as in Definition 3.1, we write the n numbers underneath in a row.) Fix an $m \times n$-matrix A. Then the product Av, which is an column vector with m entries, is an element in $\mathbf{R}^m$. We now regard this matrix A as fixed, and consider the vector v as a variable. In other words, we consider the function (or map)

$$\mathbf{R}^n \to \mathbf{R}^m, v \mapsto Av.$$

Matrix multiplication has the following basic, but crucial property.

Proposition 4.19. For any $m \times n$-matrix A, the above map is linear.

Proof. We prove this in the case $m = n = 2$ using (4.12). (The case of general m and n is just notationally more involved, but otherwise the same.) Let $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, $v' = \begin{pmatrix} v'_1 \\ v'_2 \end{pmatrix}$. Then

$$\begin{aligned} Av + Av' &= \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix} + \begin{pmatrix} a_{11}v'_1 + a_{12}v'_2 \\ a_{21}v'_1 + a_{22}v'_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}(v_1 + v'_1) + a_{12}(v_2 + v'_2) \\ a_{21}(v_1 + v'_1) + a_{22}(v_2 + v'_2) \end{pmatrix} \\ &= A \begin{pmatrix} v_1 + v_2 \\ v'_1 + v'_2 \end{pmatrix} \\ &= A(v + v'). \end{aligned}$$

Likewise, one checks (4.3), i.e., that for $a \in R$,

$$\begin{aligned} A(av) &= A \begin{pmatrix} av_1 \\ av_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}av_1 + a_{12}av_2 \\ a_{21}av_1 + a_{22}av_2 \end{pmatrix} \\ &= a \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix} \\ &= aAv \\ &= a(Av). \end{aligned}$$

4.3 Outlook: current research

Since matrix multiplication is such a key asset, it is of great interest to perform this process as efficiently as possible. Given two 2×2-matrices A and B, the computation of AB by just following the definition takes 8 multiplications, namely

$$a_{ie}b_{ej}$$

for each of the indices i, j, e being either 1 or 2. In the 1960's an algorithm (https://en.wikipedia.org/wiki/Strassen_algorithm) was found that only requires 7 multiplications. By applying that algorithm iteratively for larger matrices, this gives a decidedly better algorithm. Current research is using methods of artificial intelligence to try and come up with similar methods for 3×3- and other matrices. Check out this interesting lay-accessible article on recent trends: https://www.quantamagazine.org/ai-reveals-new-possibilities-in-matrix-multiplicati

4.4 Kernel and image of a linear map

The kernel and the image of a linear map are an important measure how, roughly speaking, interesting this map is. E.g., the zero map $\mathbf{R}^2 \to \mathbf{R}^2$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is certainly very boring in the sense that it only produces the zero vector in $\mathbf{R}^2$. By contrast, say, a rotation (by a fixed angle r) in $\mathbf{R}^2$ is more interesting, since any point in $\mathbf{R}^2$ can be obtained from another point by rotating by that angle r.

In order to introduce kernel and image, we need the following general notions related to maps between sets.

Definition 4.20. Let $f : X \to Y$ be a function between two sets.

- The *preimage* of some element $y \in Y$ is

$$f^{-1}(y) := \{x \in X \mid f(x) = y\} \ (\subset X).$$

- The *image* of f is defined as

$$\mathrm{im}\ (f) := f(X) := \{f(x) \mid x \in X\} \ (\subset Y).$$

- f is called *injective* (or *one-to-one*) if for each y, the preimage $f^{-1}(y)$ contains *at most* one element.
- f is called *surjective* (or *onto*) if for each y, $f^{-1}(y)$ contains *at least* one element. Equivalently, f is surjective if $\mathrm{im}\ (f) = Y$.
- f is called *bijective* if it is both injective and surjective. In other words, if for each $y \in Y$, $f^{-1}(y)$ contains exactly one element.

Example 4.21. While in the applications below, we will often consider X and Y to be vector spaces, Definition 4.20 applies to maps between arbitrary sets. For example, consider a group of n people $\{P_1, \dots, P_n\}$. Consider the function

$$m : \{P_1, \dots, P_n\} \to \{1, 2, \dots, 12\}$$

that assigns to each person their month of birth. This function is surjective if for each month, one of the persons is born in that month. It is injective, if in each month only one birthday party is happening. It is bijective if both conditions are true, i.e., in every month there is exactly one birthday party (for one of the persons).

In the example above, the map m can only be bijective if $n = 12$, i.e., if the size of the two sets is the same. For linear maps (between vector spaces) we want to articulate a similar idea, but simply saying that the size of the vector spaces are the same is insufficient, since $\mathbf{R}$, $\mathbf{R}^2$, $\mathbf{R}^3$ etc. all have infinitely many elements. Rather, we will see in Corollary 4.28 that the dimension of a vector space is the correct notion of size.

Definition 4.22. Let $f : V \to W$ be a linear map. The *kernel* of f is defined as

$$\begin{aligned}\ker(f) &:= f^{-1}(0_W) \\ &= \{v \in V \mid f(v) = 0_W\}.\end{aligned}$$

Note that $\ker(f) \subset V$ and $\operatorname{im}(f) \subset W$. In fact, these are not just arbitrary subsets:

Proposition 4.23. For a *linear* map $f : V \to W$, $\ker f$ is a sub*space* of V, while $\operatorname{im} f$ is a sub*space* of W.

Proof. We only check the conditions in Definition 3.17 for the kernel. (The case of the image is similar.)

- $0_V \in \ker f$: this means that $f(0_V) = 0_W$, which holds by Remark 4.4.
- For $v, v' \in \ker f$ we check $v + v' \in \ker f$: this means $f(v + v') = 0_W$. Indeed, using that f is linear we have
$$f(v + v') = f(v) + f(v') = 0_W + 0_W = 0_W.$$
- For $v \in \ker f$ and $a \in \mathbf{R}$, we check $av \in \ker f$: as before, using the linearity of f, we have
$$f(av) = af(v) = a \cdot 0_W = 0_W.$$

Example 4.24. We consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \end{pmatrix}$$

and the associated linear map

$$f : \mathbf{R}^3 \to \mathbf{R}^2, v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto Av = \begin{pmatrix} x + 2y \\ 2x + 4y \end{pmatrix}.$$

The kernel of f consists of vectors v such that

$$\begin{aligned} x + 2y &= 0 \\ 2x + 4y &= 0. \end{aligned}$$

This tells us that the kernel of f, or equivalently the solutions of this system (in the unknowns x, y and z!), is

$$\ker f = \left\{ \begin{pmatrix} -2y \\ y \\ z \end{pmatrix} \in \mathbf{R}^3 \mid y, z \in \mathbf{R} \right\}.$$

A basis of $\ker f$ is given by the two vectors $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

The image of f consists of all vectors of the form

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} x + 2y \\ 2x + 4y \end{pmatrix},$$

with arbitrary $x, y \in \mathbf{R}$. (Also, the z is arbitrary, but it does not show up in f.) This means that $v_2 = 2v_1$, and v_1 is an arbitrary real number. Thus

$$\text{im } f = \left\{ \begin{pmatrix} v_1 \\ 2v_1 \end{pmatrix} \mid v_1 \in \mathbf{R} \right\} \subset \mathbf{R}^2.$$

A basis of im f is thus given by the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

Our goal below is to develop an algorithmic method that determine bases of $\ker f$, im f. For now, just observe that in the example above

$$\dim(\ker f) + \dim(\text{im } f) = 2 + 1 = 3 = \dim \mathbf{R}^3.$$

This is an example of the rank-nullity theorem (Theorem 4.26) below.

Injectivity of *linear* maps can be measured in terms of the kernel:

Lemma 4.25. Let $f : V \to W$ be a linear map. Then the following are equivalent (i.e., one condition holds if and only if the other holds):

(1) f is injective,

(2) $\ker f = \{0_V\}$.

Proof. Suppose f is injective, we prove $\ker f = \{0\}$. Since $f(0) = 0$ by linearity (Remark 4.4), we have $0 \in \ker f$. If $v \in \ker f$, then $f(v) = 0_W$, so both v and 0_V are in the preimage of 0_W. By the injectivity of f, this forces $v = 0$.

Conversely, suppose $\ker f = 0$. Suppose two vectors $v, v' \in V$ are in the preimage of some $w \in W$, i.e., $f(v) = f(v') = w$. Then, by linearity of f

$$f(v - v') = f(v + (-1)v') = f(v) + (-1)f(v') = f(v) - f(v') = 0.$$

Thus, $v - v' \in \ker f$, which means by assumption that $v - v' = 0$. That is: $v = v'$. Therefore f is injective. □

Theorem 4.26. (*Rank-nullity theorem*) Let $f : V \to W$ be a map between (finite-dimensional) vector spaces. Then

$$\dim(\ker f) + \dim(\text{im } f) = \dim V.$$

The *rank* of f is defined to be

$$\text{rk} f := \dim(\text{im } f),$$

while the *nullity* of f is defined to be $\dim(\ker f)$.

A proof of this theorem appears in any linear algebra textbook, e.g. [Nic95, Theorem 7.2.4]. As a remark on the proof, we note that one can prove the following fact, which is very useful in its own right.

Theorem 4.27. Let $f : V \to W$ be a linear map. Let

$$v_1, \dots, v_r, v_{r+1}, \dots v_n$$

be a basis of V such that

$$v_1, \dots, v_r$$

is a basis of $\ker f$. Then $f(v_{r+1}), \dots, f(v_n)$ is a basis of $\text{im } f$.

The following facts are immediate consequences of the rank-nullity theorem.

Corollary 4.28. Let $f : V \to W$ be a linear map between finite-dimensional vector spaces.

(1) If f is injective then $\dim V \leqslant \dim W$ (since then $\ker f = \{0\}$, i.e., $\dim \ker f = 0$).

(2) If f is surjective then $\dim V \geqslant \dim W$ (since them $\operatorname{im} f = W$, so $\dim \operatorname{im} f = \dim W$).

(3) If f is bijective then $\dim V = \dim W$.

(4) The preceding three statements can in general not be reversed: if, say, $\dim V \leqslant \dim W$, f need not be injective. For example the zero map $V \to W$, $v \mapsto 0$ is *never* injective if $V \neq \{0\}$.

(5) Suppose in addition that $\dim V = \dim W$. In this case f is injective precisely if f is surjective. (If f is injective, then $\dim \operatorname{im} f = \dim V = \dim W$, so that $\operatorname{im} f = W$ by Theorem 3.67(3). Similarly, if f is surjective, then $\dim \operatorname{im} f = \dim W = \dim V$, so $\dim \ker f = 0$, so that $\ker f = \{0\}$.)

An important case of this theorem is where $f : \mathbf{R}^n \to \mathbf{R}^m$ is the linear map given by multiplication with a fixed $m \times n$-matrix A. We call the *rank* of A, resp. the *nullity* the rank, resp. nullity of that linear map. The rank is denoted by $\operatorname{rk} A$. These are two highly important numbers associated to a matrix, so we want to have a device for computing them. This is based on the following computation: recall from Example 3.59 the standard basis vectors

$$e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots 0), \dots, e_n = (0, \dots, 0, 1) \in \mathbf{R}^n.$$

We will in the sequel write them as column vectors, so $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ etc. Then we have

$$f(e_i) = Ae_i = \begin{pmatrix} a_{11} \cdot 0 + \dots + a_{1i} \cdot 1 + \dots + a_{1n} \cdot 0 \\ \vdots \\ a_{m1} \cdot 0 + \dots + a_{mi} \cdot 1 + \dots + a_{mn} \cdot 0 \end{pmatrix} = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}. \tag{4.29}$$

In other words, the product Ae_i is precisely the i-th column of the matrix A!

Since any vector $v \in \mathbf{R}^n$ is a linear combination of the e_i, we have, for appropriate $b_1, \dots, b_n \in \mathbf{R}$

$$f(v) = f(\sum_{i=1}^{n} b_i e_i) = \sum_{i=1}^{n} b_i f(e_i).$$

Thus, $f(v)$ is a linear combination of the columns of A. This proves the following statement:

Proposition 4.30. Let A be an $m \times n$-matrix and $f : \mathbf{R}^n \to \mathbf{R}^m$ the linear map given by multiplication with A. We write

$$A = (c_1 \ c_2 \ \dots c_n),$$

i.e., the $c_i (\in \mathbf{R}^m)$ is the i-th column of A. Then

$$\operatorname{im} f = L(c_1, \dots, c_n).$$

This subspace of $\mathbf{R}^m$ is also called the *column space* of A.

Definition 4.31. The *row space* of A is the subspace of $\mathbf{R}^n$ spanned by the rows of the matrix A.

We can compute the rank of A, i.e., the $\dim \operatorname{im} f$, as follows:

Proposition 4.32. Let A be an $m \times n$-matrix. Suppose B is a (possibly non-reduced) row-echelon matrix obtained from A by means of elementary row operations (Definition 2.29).

(1) Then the non-zero rows of B form a basis of the row space of A.

(2) If the leading ones of B lie in the columns $j_1, \dots, j_r$, then these columns of A form a basis of the column space of A.

(3) The following numbers are all the same: a) $\operatorname{rk} A$, b) the dimension of the column space, c) the dimension of the row space of A, d) the number of leading ones in B.

(4) The nullity of A equals n (the number of columns of A) minus any of the quantities in the previous point.

Proof. Parts (1) and (2) can be proven by showing that the row and column space of A do not change when one performs an elementary row operation to A. We skip this part of the proof (e.g., see [Nic95, Lemma 5.4.1] for a proof).

(3) follows from the first two: by definition, $\operatorname{rk} A = \dim \operatorname{im} f$ equals, by Proposition 4.30, the dimension of the column space. By the second statement, this is equal to the number of leading ones in B. Since B is a row-echelon matrix, this is also the number of non-zero rows, i.e., by the first statement, the dimension of the row space.

Finally, (4) is a consequence of the rank-nullity-theorem. □

Example 4.33. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix}$$

and the linear map

$$f : \mathbf{R}^4 \to \mathbf{R}^3, v \mapsto Av.$$

The row space is the subspace of $\mathbf{R}^4$ spanned by the vectors $(1\ 2\ 2\ -1)$ etc., while the column space is the subspace of $\mathbf{R}^3$ spanned by the vectors $\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$ etc. We compute a basis of these two spaces as follows:

$$A \rightsquigarrow \begin{pmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & -1 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus, the vectors $(1, 2, 2, -1)$ and $(0, 0, 1, -3)$ form a basis of the row space. In particular, its dimension is two. The first and third row of A form a basis of the column space of A, i.e.,

$$\operatorname{im}\ f = L(\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}).$$

Thus

$$\dim \operatorname{im}\ f = \operatorname{rk} f = \operatorname{rk} A = 2.$$

According to the rank-nullity theorem (Theorem 4.26),

$$\dim \ker f = \dim \mathbf{R}^4 - \dim \operatorname{im}\ f = 4 - 2 = 2,$$

(i.e., the nullity of f or of A is 2). In order to determine a basis of $\ker f$, we denote the coordinates in $\mathbf{R}^4$ by $x_1, \dots, x_4$. Then, according to Gaussian elimination, the variables x_2 and x_4 are free variables, and $x_3 = 3x_4$ from the second row above, and then $x_1 = -2x_2 - 2x_3 + x_4 = -2x_2 - 5x_4$. Thus a basis of $\ker f$ is given by the vectors

$$\begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 3 \\ 1 \end{pmatrix}.$$

This reconfirms that $\dim \ker f = 2$.

Remark 4.34. Even though the dimension of the row space and the column space are the same, these vector spaces themselves are *not* the same. In fact, they are not even comparable, given that the row space is a subspace of $\mathbf{R}^n$, while the column space is a subspace of $\mathbf{R}^m$.

Here is another consequence of the rank-nullity theorem.

Theorem 4.35. (stated above in Theorem 3.73) Suppose $A, B \subset V$ are two subspaces of a vector space. Then

$$\dim(A \cap B) + \dim(A + B) = \dim A + \dim B.$$

Proof. The map

$$f : A \oplus B \to V, (a, b) \mapsto a - b$$

is linear. Since for every $b \in B$ also $b' := -b$ is contained in B, the image of this map is $A + B$. The kernel of f consists of those vectors $(a, b) \in A \oplus B$ such that $a - b = 0$, i.e., $a = b$. This means that $a \in A \cap B$. Therefore, the rank nullity theorem and Example 3.64 tell us

$$\dim(A \cap B) + \dim(A + B) = \dim \ker f + \dim \operatorname{im} f = \dim(A \oplus B) = \dim A + \dim B. \blacksquare$$

4.5 Revisiting linear systems

In this section, we apply our findings from above to the problem of solving a linear system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Throughout, let $A = (a_{ij})$ be the $m \times n$-matrix formed by the coefficients of that linear system. Recall that the vector

$$b := \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

is called the vector of *constants*. We will also consider the linear map (Proposition 4.19)

$$f : \mathbf{R}^n \to \mathbf{R}^m, v \mapsto Av.$$

Theorem 4.36. (1) Suppose momentarily that $b_1 = \cdots = b_m = 0$, so the above system is homogeneous. In this case the solution set equals $\ker f$, which in particular is a sub*space* of $\mathbf{R}^n$.

(2) For arbitrary b, the system above has (at least) one solution if the vector b lies in the image of f. (Note that the vector is $\mathbf{R}^m$, and $\operatorname{im} f \subset \mathbf{R}^m$.) If $r = (r_1, \dots, r_n)$ is any such solution, then the solution set consists precisely of the vectors of the form

$$r + \ker f := \{r + v, \text{ where } v \in \ker f\}.$$

Proof. Recall from Observation 4.11 that

$$f^{-1}(b) = \{r \in \mathbf{R}^n \mid Ar = b\}$$

consists precisely of the solutions of the system above.

Therefore, the first statement is clear: $\ker f = f^{-1}(0)$ are the solutions of the homogeneous system. Also, the (non-homogeneous) system has a solution precisely if $f^{-1}(b)$ is non-empty, i.e., if $b \in \operatorname{im} f$. For the last statement: we show both implications:

- if $s = (s_1, \ldots, s_n)$ is a solution, then we get

$$f(s - r) = f(s) - f(r)$$

  since f is linear. Since r is some solution of the system, we have $f(r) = b$, and also $f(s) = b$. This implies $v := s - r \in \ker f$, i.e., $s = r + v$.

- Conversely, consider a vector of the form $r + v$, with $v \in \ker f$. Then

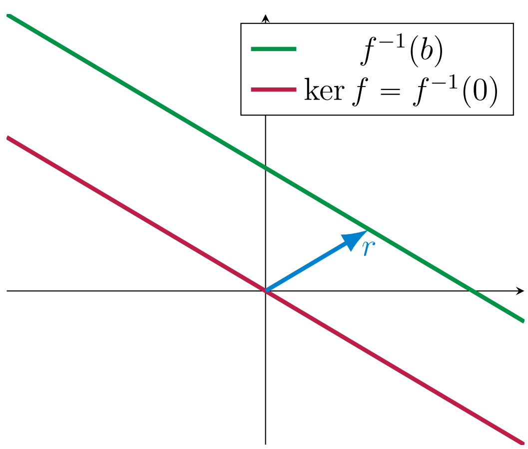
$$f(r + v) = f(r) + f(v) = b + 0 = b.$$

  Thus $r + v$ is also a solution of the system. □

Remark 4.37. The solution set $r + \ker f$ of a non-homogeneous system is *never* a subspace: indeed any subspace contains the zero vector, but if that is a solution we get

$$b = A0 = 0.$$

Instead, the solution set of the system with a non-zero vector b, i.e., $f^{-1}(b)$ is a translation of $\ker f$, as is



Example 4.38. Consider the linear system (in the unknowns x, y, z)

$$\begin{aligned} x + 3y + 5z &= 7 \\ 3x + 9y + 10z &= 11 \\ 2x + 9y + 12z &= 10. \end{aligned}$$

The pertinent 3×3-matrix built out of the coefficients is

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 9 & 10 \\ 2 & 9 & 12 \end{pmatrix}.$$

As above, we write $f : \mathbf{R}^3 \to \mathbf{R}^3, v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto Av$ for the associated linear map.

We compute its rank by bringing it into row-echelon form:

$$A \rightsquigarrow \begin{pmatrix} 1 & 3 & 5 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 3 & 5 \\ 0 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 3 & 5 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \overset{\text{swap}}{\rightsquigarrow} \begin{pmatrix} 1 & 3 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This matrix has 3 leading ones, hence its rank is 3. Thus, f is surjective. By the rank-nullity theorem we have

$$\dim \ker f = \dim \mathbf{R}^3 - \dim \operatorname{im} \, f = 3 - 3 = 0.$$

Therefore, f is injective (Lemma 4.25). (Alternatively, we may use Corollary 4.28(5) directly to see f is injective.) Thus, f is bijective. This means that for *any* vector of constants, such as the above $\begin{pmatrix} 7 \\ 11 \\ 10 \end{pmatrix}$, there is precisely one solution of the linear system. This solution can be determined via Method 2.32, but we will omit this computation here because we will later develop a more comprehensive method, namely by using the *inverse* A^{-1}, to obtain these solutions.

4.6 Linear maps defined on basis vectors

An arbitrary map

$$f : V \to W$$

encodes a lot of information: one needs to specify $f(v)$ for *every* $v \in V$. For *linear* maps, this simplifies drastically:

Proposition 4.39. Let $v_1, \dots, v_n$ be a basis of a vector space V. Let W be another vector space and $w_1, \dots, w_n$ arbitrary vectors (they need not be linearly independent, or span W etc.) Then there is a *unique* linear map $f : V \to W$ such that

$$f(v_i) = w_i. \tag{4.40}$$

Proof. Recall Proposition 3.61: given a basis $v_1, \dots, v_n$ of a vector space, any vector $v \in V$ can be *uniquely* expressed as a linear combination

$$v = b_1 v_1 + \dots + b_n b_n = \sum_{i=1}^n b_i v_i,$$

i.e., we can express v in such a form and the real numbers b_i are uniquely determined by v. Moreover, we can think of these numbers $b_1, \dots, b_n$ as the coordinates of v (with respect to our coordinate system given by the basis). Namely, given another vector $v' = \sum_{i=1}^n b'_i v_i$ and some $a \in \mathbf{R}$, we have

$$v + v' = \sum_{i=1}^n (b_i + b'_i) v_i$$
$$av = \sum_{i=1}^n (ab_i) v_i.$$

Now, given $v \in V$, we *define*

$$f(v) := \sum_{i=1}^n b_i w_i. \tag{4.41}$$

In particular, for $v = v_i$, this satisfies $f(v_i) = w_i$. The map f is linear; this follows from the preceding discussion.

Conversely, if a linear map f satisfies $f(v_i) = w_i$, for v as above, it necessarily satisfies

$$f(v) = f(\sum_{i=1}^n b_i v_i) = \sum_{i=1}^n b_i f(v_i) = \sum_{i=1}^n b_i w_i.$$

So, the map defined in (4.41) is the only linear map satisfying (4.40). □

Example 4.42. We consider $V = \mathbf{R}^3$, with the basis

$$v_1 = e_1 = (1, 0, 0), v_2 = e_2 = (0, 1, 0), v_3 = (0, 1, -1).$$

(Note that e_1, e_2 are part of the standard basis of $\mathbf{R}^3$.) According to Proposition 4.39, there is a unique linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ such that

$$f(v_1) = (2, -1, 0), f(v_2) = (1, -1, 1), f(v_3) = (0, 2, 2).$$

We determine $f(e_3)$, where $e_3 = (0, 0, 1)$ is the third standard basis vector. We have

$$e_3 = v_2 - v_3.$$

Thus

$$f(e_3) = f(v_2 - v_3) = f(v_2) - f(v_3) = (1, -1, 1) - (0, 2, 2) = (1, -3, -1).$$

Thus, with respect to the *standard basis* e_1, e_2, e_3 (which is distinct from the one above!), the matrix of f is given by

$$A = \begin{pmatrix} 2 & 1 & 1 \\ -1 & -1 & -3 \\ 0 & 1 & -1 \end{pmatrix}.$$

That is, f agrees with the map

$$f : \mathbf{R}^3 \to \mathbf{R}^3, v \mapsto Av.$$

4.7 Matrices associated to linear maps

In Proposition 4.19, we associated a linear map $\mathbf{R}^n \to \mathbf{R}^m$ to an $m \times n$-matrix. In this section, we will reverse this process: we will begin with a linear map and associate to it a matrix.

Proposition 4.43. Let V, W be two vector spaces with bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$, respectively. Let finally $f : V \to W$ be a linear map. Then there is a unique $m \times n$-matrix $A = (a_{ij})$, called the *matrix associated to the linear map* f *with respect to the given bases*, such that

$$f(v_i) = \sum_{j=1}^{m} a_{ji} w_j.$$

We denote this matrix by

$$\mathrm{M}_{f,\underline{v},\underline{w}} := \mathrm{M}_{f,\{v_1,\dots,v_n\},\{w_1,\dots,w_m\}} := (a_{ij}),$$

where for brevity $\underline{v} := \{v_1, \dots, v_n\}$ and $\underline{w} := \{w_1, \dots, w_m\}$.

For a general vector $v = \sum_{i=1}^n b_i v_i$, we have

$$f(v) = \sum_{i=1}^n \sum_{j=1}^m b_i a_{ji} w_j.$$

Proof. We apply the above fact (Proposition 3.61) to $f(v_i) \in W$ (and the basis $w_1, \dots, w_m$), and see immediately that a unique expression of $f(v_i)$ as claimed exists.

We now compute $f(v)$:

$$\begin{aligned} f(v) &= f(\sum_{i=1}^n b_i v_i) \\ &= \sum_{i=1}^n b_i f(v_i) && \text{since } f \text{ is linear} \\ &= \sum_{i=1}^n b_i \sum_{j=1}^m a_{ji} w_j \\ &= \sum_{i=1}^n \sum_{j=1}^m b_i a_{ji} w_j. \end{aligned}$$

Example 4.44. We continue Example 4.42. The vectors $w_1 = f(v_1) = (2, -1, 0)$, $w_2 = f(v_2) = (1, -1, 1)$ and $w_3 = f(v_3) = (0, 2, 2)$ form a basis of $\mathbf{R}^3$, as one sees by computing the rank of

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 1 \\ 0 & 2 & 2 \end{pmatrix},$$

which is three. We can therefore apply Proposition 4.43 to the bases v_1, v_2, v_3 and w_1, w_2, w_3. The matrix is then

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}!$$

To see this, note for example the second row says

$$f(e_2) = 0w_1 + 1w_2 + 0w_3,$$

which is true.

If, by contrast, we consider the standard basis e_1, e_2, e_3 of $V = \mathbf{R}^3$ (and still w_1, w_2, w_3 in $W = \mathbf{R}^3$), then the matrix reads

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

For example, the third column of this matrix expresses the identity

$$f(e_3) = a_{13}w_1 + a_{23}w_2 + a_{33}w_3 = w_2 - w_3,$$

which we computed above.

This in particular shows that the matrix A depends (not only on f but also on) the choice of the bases of V and W!

Example 4.45. We consider the rotation matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, cf. Example 4.17, and consider the associated linear map

$$f : \mathbf{R}^2 \to \mathbf{R}^2, v = \begin{pmatrix} x \\ y \end{pmatrix} \mapsto Av = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

We consider the basis $\underline{v}$ of $\mathbf{R}^2$ consisting of $v_1 = (1, 0)$ and $v_2 = (1, 1)$. We compute the basis of f with respect to $\underline{v}$ on the domain $\mathbf{R}^2$, and the standard basis on the codomain $\mathbf{R}^2$. In order to compute it, we need to express v_1 and v_2 in terms of the standard basis, which is straightforward:

$$\begin{aligned} v_1 &= 1 \cdot e_1 + 0 \cdot e_2 \\ v_2 &= 1 \cdot e_1 + 1 \cdot e_2. \end{aligned}$$

The linearity of f implies

$$f(v_1) = 1 \cdot f(e_1) + 0 \cdot f(e_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \underbrace{0}_{a_{11}} \cdot e_1 + \underbrace{1}_{a_{21}} \cdot e_2.$$

$$f(v_2) = 1 \cdot f(e_1) + 1 \cdot f(e_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \underbrace{-1}_{a_{12}} \cdot e_1 + \underbrace{1}_{a_{22}} \cdot e_2.$$

.

Thus, the matrix of f with respect to afore-mentioned bases is

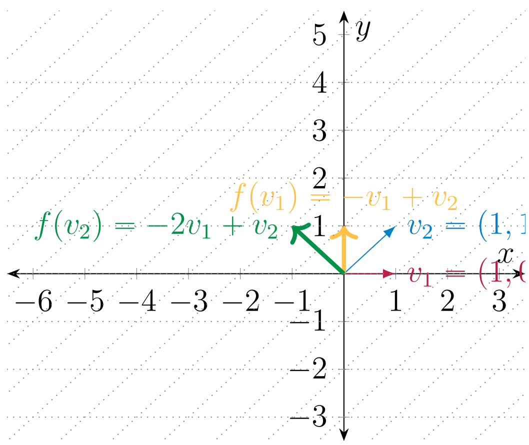
$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}.$$

We additionally compute the matrix of f with respect to the basis $\underline{v}$ both on the domain and on the codomain. To this end, we need to express the vectors $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ as linear combinations of v_1 and v_2. We have

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -v_1 + v_2$$
$$\begin{pmatrix} -1 \\ 1 \end{pmatrix} = -2v_1 + v_2.$$

Thus, the matrix of f with respect to the basis $\underline{v}$ on both the domain and the codomain is

$$\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$



4.8 Composing linear maps and multiplying matrices

The following lemma, while simple to prove, is of fundamental importance:

Definition and Lemma 4.46. Let $f : U \to V$ and $g : V \to W$ be two linear maps between three vector spaces U, V and W. Then the *composition* of g and f is the map defined as

$$g \circ f : U \to W, u \mapsto g(f(u)).$$

This map is again linear.

Proof. We check the two conditions in Definition 4.1: for $u, u' \in U$ and $a \in \mathbf{R}$, we have, using the linearity of f and g:

$$\begin{aligned}(g \circ f)(u + u') &= g(f(u + u')) \\ &= g(f(u) + f(u')) \\ &= g(f(u)) + g(f(u')) \\ &= (g \circ f)(u) + (g \circ f)(u') \\ (g \circ f)(au) &= g(f(au)) \\ &= g(af(u)) \\ &= ag(f(u)) \\ &= a(g \circ f)(u).\end{aligned}$$

Example 4.47. The maps $f : \mathbf{R}^2 \to \mathbf{R}$, $(x, y) \mapsto x$ and $g : \mathbf{R} \to \mathbf{R}^3$, $x \mapsto (x, 0, x)$ are both linear. The composition $g \circ f$ is the map

$$g \circ f, (x, y) \mapsto g(f(x, y)) = g(x) = (x, 0, x).$$

We may also consider $h : \mathbf{R} \to \mathbf{R}^2$, $x \mapsto (x, x)$. Then the composite

$$h \circ f, (x, y) \mapsto h(f(x, y)) = h(x) = (x, x).$$

The other composite is also defined, it is

$$f \circ h : \mathbf{R} \to \mathbf{R}, x \mapsto f(h(x)) = f(x, x) = x.$$

(By comparison, the composition $f \circ g$ is not defined, since g takes values in $\mathbf{R}^3$, but f is defined on $\mathbf{R}^2$.)

We now relate this composition of abstract maps to something more concrete, the product of matrices.

Definition 4.48. If $A = (a_{ij})$ is a $m \times n$-matrix and $B = (b_{ij})$ is an $n \times k$-matrix, then the *product* AB (also sometimes denoted by $A \cdot B$) is the $m \times k$-matrix whose entry in the i-th row and j-th column is the following (see §A for the sum notation $\sum$):

$$\sum_{e=1}^{n} a_{ie} b_{ej} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}.$$

In other "words"

$$AB := (\sum_{e=1}^{n} a_{ie} b_{ej}).$$

I.e., one picks the i-th row of A and the j-th column of B; one traverses these and multiplies the corresponding entries together one by one and finally adds up these products.

Example 4.49.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 6 & -2 \end{pmatrix} = \begin{pmatrix} 1\cdot(-1)+2\cdot 6 & 1\cdot 0+2\cdot(-2) \\ 3\cdot(-1)+4\cdot 6 & 3\cdot 0+4\cdot(-2) \end{pmatrix}$$
$$= \begin{pmatrix} 11 & -4 \\ 21 & -8 \end{pmatrix},$$

$$\begin{pmatrix} 1 & -1 & 2 \\ 1 & 3 & -2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{pmatrix} =$$
$$=$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & -1 & 2 \\ 1 & 3 & -2 \end{pmatrix} =$$
$$=$$

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} =$$
$$=$$

Note that the second product is a 2×2-matrix while the product of the *same* matrices in the other order is a 3×3-matrix!

The product AB is *only* defined if the number of columns of A is the same as the number of rows of B. For example,

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 2 & 3 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}$$

is *not* defined, i.e., it is a meaningless expression.

Remark 4.50. In the case when B is a column vector with n entries, we can regard it as an $n \times 1$-matrix. In this case the product AB defined in Definition 4.48 is an $m \times 1$-matrix, which agrees with the column vector AB as defined in Definition 4.9, so the product considered now is a generalization of that previous construction. In general, if B is an $n \times k$-matrix, we can write it as

$$B = (b_1 \ b_2 \ \dots \ b_n),$$

where the $b_1, \dots, b_n$ are the columns of B. Then

$$AB = (Ab_1 \ Ab_2 \ \dots \ Ab_n).$$

In Proposition 4.19, we associated to an $m \times n$-matrix A a linear map

$$f : \mathbf{R}^n \to \mathbf{R}^m, v \mapsto Av.$$

Let us also be given an $n \times l$-matrix B, to which we can assign the linear map

$$g : \mathbf{R}^l \to \mathbf{R}^n, u \mapsto Bu.$$

Proposition 4.51. In the above situation, the compositition $f \circ g : \mathbf{R}^l \to \mathbf{R}^n$ is the map given by multiplication by the matrix AB, i.e., the linear map

$$u \mapsto (AB)u.$$

Proof. Let us write $C = AB$ for the product of A and B. It is an $m \times l$-matrix. If we write $C = (c_{ij})$, we have

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}. \tag{4.52}$$

We have to compare two linear maps, $\mathbf{R}^l \to \mathbf{R}^n$, namely $f \circ g$ and $u \mapsto Cu = (AB)u$. According to Proposition 4.39, it suffices to show that these two maps give the same values when we evaluate them on some basis of $\mathbf{R}^n$, for which we take the standard basis $e_1, \dots, e_n$. As was noted in (4.29), the product Ce_i is precisely the i-th column of C. That is,

$$Ce_i = \begin{pmatrix} c_{1i} \\ \vdots \\ c_{mi} \end{pmatrix} = c_{1i}e_1 + \dots + c_{mi}e_m = \sum_{s=1}^{m} c_{si}e_s = \sum_{s=1}^{m} \sum_{r=1}^{n} a_{sr}b_{ri}e_s.$$

Similarly,

$$f(e_i) = Ae_i = \sum_{s=1}^{m} a_{si}e_s$$

and

$$g(e_i) = Be_i = \sum_{r=1}^{n} b_{ri}e_r.$$

Here, as usual, $e_1, \dots$ denotes the standard basis vectors of $\mathbf{R}^n$, $\mathbf{R}^m$ and $\mathbf{R}^l$. We now compute

$$\begin{aligned}
(f \circ g)(e_i) &= f(g(e_i)) \\
&= f(\sum_{r=1}^n b_{ri} e_r) \\
&= \sum_{r=1}^n b_{ri} f(e_r) && (f \text{ is linear}) \\
&= \sum_{r=1}^n b_{ri} \sum_{s=1}^m a_{sr} e_s \\
&= \sum_{r=1}^n \sum_{s=1}^m b_{ri} a_{sr} e_s \\
&= \sum_{s=1}^m \sum_{r=1}^n a_{sr} b_{ri} e_s \\
&= \sum_{s=1}^m c_{si} e_s. && \text{by (4.52).}
\end{aligned}$$

With similar arguments, one proves the following:

Proposition 4.53. Let $f : U \to V$ and $g : V \to W$ be two linear maps, and let $u_1, \dots, u_l$, $v_1, \dots, v_m$ and $w_1, \dots, w_n$ be bases of these vector spaces. Finally, let A be the matrix of f with respect to these bases (of U and V) and B the matrix of g with respect to these bases (of V and W). Then BA is the matrix of $g \circ f$ with respect to the bases (of U and W).

4.8.1 Properties of matrix multiplication

Dependence on the order of factors

A key property of matrix multiplication is that the product of two matrices depends on the *order* of the factors.

Warning 4.54. For two $n \times n$-matrices A and B, their product depends on the order of the two matrices. In other words, in general

$$AB \neq BA!$$

Mark these words! It is a common misconception among linear algebra-learners to think that AB would (always) be equal to BA.

Example 4.55. Examples are not hard to come by:

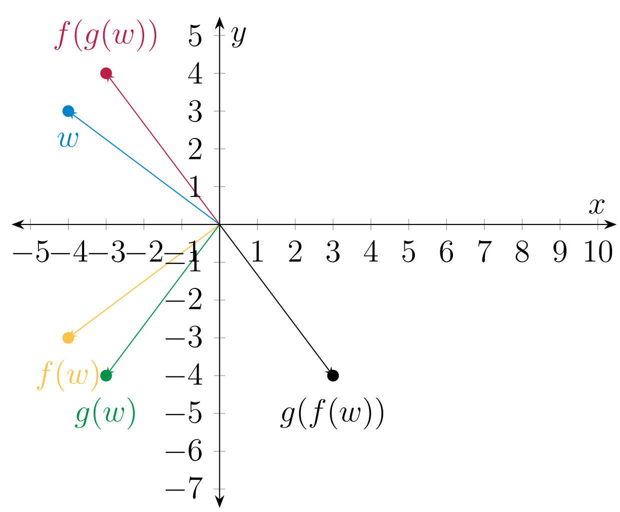
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & \\ & \end{pmatrix}$$
$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \\ & \end{pmatrix}$$

So that
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}!$$

Remark 4.56. The phenomenon $AB \neq BA$ may be best understood in the light of composition of (linear) maps: if $f : \mathbf{R}^n \to \mathbf{R}^n$ and $g : \mathbf{R}^n \to \mathbf{R}^n$ is another linear map, then in general we have

$$g \circ f \neq f \circ g.$$

To take a concrete example, consider the linear map $f : \mathbf{R}^2 \to \mathbf{R}^2$ given by reflecting along the x-axis, and $g : \mathbf{R}^2 \to \mathbf{R}^2$ the linear map given by rotating counter-clockwise (around the origin) by 90°.



Let us conclude this discussion by noting that this issue is not specific to linear algebra, but is a common phenomenon in daily life: there is (often) no reason to expect that doing (the same) two actions in different order give the same result:

- You first do sports, then take a shower.
- You first take a shower, then do sports.

In the first scenario you may feel refreshed, in the second one a little sweaty...

Further properties of matrix multiplication

Definition 4.57. The *identity matrix* is the square matrix

$$\mathrm{id} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

I.e., it is a square matrix whose entries on the "north-west – south-east" diagonal (which is called the *main diagonal*) are all 1, and the remaining entries are zero. If it is important to specify the size, one also writes id_n.

Example 4.58. If $n = 1$, then id_1 is just the 1×1-matrix whose only entry is 1. $\mathrm{id}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The first two identities in the next lemma assert that the identity matrix takes the role of the number 1 when it comes to multiplying matrices.

Lemma 4.59. Matrix multiplication satisfies the following identities, where A, B and C are matrices (of a size such that the products and sums below are defined), and $r \in \mathbf{R}$:

$$\begin{aligned} \mathrm{id}A &= A \\ A\mathrm{id} &= A \\ A(B + C) &= AB + AC && \text{(distributivity)} \\ (A + B)C &= AC + BC \\ (AB)C &= A(BC) && \text{(associativity)} \\ r(AB) &= (rA)B = A(rB) && \text{(matrix vs. scalar multiplication)} \end{aligned}$$

Proof. These identities follow from similar identities for the multiplication and addition of real numbers.

To illustrate the principle, we consider the first distributivity law above. Let $A = (a_{ij})$ be an $m \times n$-matrix and B, C two $n \times k$-matrices, $B = (b_{ij})$ and $C = (c_{ij})$. Then $B + C = (b_{ij} + c_{ij})$ so that

$$\begin{aligned} A(B+C) &= (\sum_{e=1}^{n} a_{ie}(b_{ej} + c_{ej})) \\ &\stackrel{!}{=} (\sum_{e=1}^{n} a_{ie}b_{ej} + a_{ie}c_{ej}) \\ &= (\sum_{e=1}^{n} a_{ie}b_{ej}) + (\sum_{e=1}^{n} a_{ie}c_{ej}) \\ &= AB + AC. \end{aligned}$$

At the equality marked ! we have used the distributivity law for real numbers, i.e., the identity $e(f+g) = ef + eg$ for any $e, f, g \in \mathbf{R}$. □

Multiplication with elementary matrices

We recast the elementary row operations of matrices (Definition 2.29) in terms of multiplication with appropriate matrices. Below, we use the (standard) convention that an "invisible" entry in a matrix is zero, e.g. $\mathrm{id}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ will be written as $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ etc.

Proposition 4.60. Let A be an $m \times n$-matrix.

(1) Let A' be the matrix obtained by interchanging the i-th and the j-th row. Then

$$A' = \underbrace{\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}}_{E^{(1)}_{i,j}} A.$$

(The first matrix is the $m \times m$-matrix obtained from id_m by exchanging the i-th and the j-th row.)

(2) Let A' be the matrix obtained by multiplying the i-th row with a real number r. Then

$$A' = \underbrace{\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & r & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}}_{E^{(2)}_{i,r}} A.$$

(The first matrix is the $m \times m$-matrix obtained from id_m by replacing the (i,i)-entry by r.)

(3) Let A' be the matrix obtained by adding the r-th multiple of the j-th row to the i-th row. Then

$$A' = \underbrace{\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}}_{E^{(3)}_{i,j,r}} A.$$

(The first matrix is the $m \times m$-matrix obtained from id_m by replacing the (i,j)-entry by r.)

Definition 4.61. The matrices $E^{(1)}_{i,j}$, $E^{(2)}_{i,r}$ and $E^{(3)}_{i,j,r}$ (for any appropriate i, j and any $r \in \mathbf{R}$, where $r \neq 0$ in $E^{(2)}_{i,r}$) appearing in the statement above are called *elementary matrices*.

Proof. This is a more cumbersome to write down precisely than to convince oneself by unwinding the definition. We check the third statement. If $B = (b_{ij})$ is the above matrix as stated, we have that $b_{ii} = 1$ and $b_{ij} = r$ and all other entries are zero. Let us write

$C = BA$, $C = (c_{ij})$. Then, by definition,

$$c_{st} = \sum_{e=1}^{m} b_{se} a_{et}.$$

We compute this sum:

- if $s \neq i$, then the only b_{se} that is non-zero is $b_{ss} = 1$, so that
$$c_{st} = b_{ss} a_{st} = a_{st}.$$
- For $s = i$, the only coefficients b_{se} that are non-zero are $b_{ss} = 1$ and $b_{sj} = r$. Thus, the sum above consists of two terms, and therefore
$$c_{st} = b_{ss} a_{st} + b_{sj} a_{jt} = a_{st} + r a_{jt}.$$

Thus the i-th row of C equals the matrix A' as in the statement above. □

4.9 Inverses

Given a linear map $f : V \to W$ it is a natural question whether the process of applying f can be undone. For example, if f encodes a counter-clockwise rotation in the plane by 60°, it can be undone by rotating clockwise by 60°. On the other hand, the linear map

$$\mathbf{R}^2 \to \mathbf{R}^2, (x, y) \mapsto (x, 0)$$

cannot be undone, since there is no way of recovering (x, y) only from x.

Definition and Lemma 4.62. Let $f : V \to W$ be a linear map. Then the following statements are equivalent (i.e., one holds precisely if the other holds):
(1) f is bijective (Definition 4.20),

(2) There is a linear map $g : W \to V$ such that

$$g \circ f = \mathrm{id}_V \text{ and } f \circ g = \mathrm{id}_W.$$

(By definition of the composition (see also §A) this means $g(f(v)) = v$ for all $v \in V$ and $f \circ g = \mathrm{id}_W$ (i.e., $f(g(w)) = w$ for all $w \in W$.)

If this is the case, we call f an *isomorphism*. In this event, the following statements hold:

- Such a map g is unique. It is also called the *inverse* of f and is denoted by $f^{-1} : W \to V$.
- $\dim V = \dim W$.

Proof. We only prove the direction (1) $\Rightarrow$ (2). By assumption f is bijective, i.e., the preimage $f^{-1}(w)$ consists of precisely one element, say $f^{-1}(w) = \{v\}$. (That is, only for that vector do we have that $f(v) = w$.) We define a map $g : W \to V$ by $g(w) := v$.

To compute $g(f(v))$ we observe that $f^{-1}(f(v)) = \{v\}$, since v is the only element of V that is mapped to $f(v)$. Thus $g(f(v)) = v$.

To compute $f(g(w))$, say that $f^{-1}(w) = \{v\}$. This means in particular that $f(v) = w$. Then $g(w) = v$ and therefore $f(g(w)) = w$.

We show that g is linear. Let $w, w' \in W$ be given. Let $v, v' \in V$ be the unique elements such that $f(v) = w$, $f(v') = w'$. By definition, this means $g(w) = v$, $g(w') = v'$. Then $w + w' = f(v + v')$, since f is linear. Thus $g(w + w') = v + v' = g(w) + g(w')$. In a similar way, one shows $g(aw) = ag(w)$ for $a \in \mathbf{R}$.

If $g' : W \to V$ is another map with $f(g'(w)) = w$, as above, then

$$f(g'(w)) = f(g(w)).$$

Since f is injective, this implies $g'(w) = g(w)$. This shows that g is unique.

The last statement holds by Corollary 4.28(3). $\square$

4.9.1 Definition and unicity of inverses

Definition 4.63. Let A be an $n \times n$-matrix. Another $n \times n$-matrix B is called an *inverse* of A if

$$AB = \mathrm{id} \text{ and } BA = \mathrm{id}.$$

If such a matrix B exists, A is called *invertible*.

Example 4.64. $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ is invertible, since $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$ is an inverse of A:

$$AB = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} =$$
$$BA = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} =$$

Not every matrix has an inverse. An 1×1-matrix A, which is just a single real number a is invertible precisely if $a \neq 0$. In this case the 1×1-matrix with entry $\frac{1}{a}$ is an inverse. For larger matrices, it is not enough to be different from zero in order to be invertible, as the following example shows.

Example 4.65. The matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

is *not* invertible. We prove this by taking an arbitrary 2×2-matrix $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and computing

$$AB = \begin{pmatrix} a + 2c & b + 2d \\ 2a + 4c & 2b + 4d \end{pmatrix}$$

Thus the condition $AB = \mathrm{id} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ amounts to four equations:

$$\begin{aligned} a + 2c &= 1 \\ b + 2d &= 0 \\ 2a + 4c &= 0 \\ 2b + 4d &= 1. \end{aligned}$$

Indeed, multiplying the first equation by 2 gives $2a + 4c = 2$, and inserting the third equation gives a contradiction:

$$0 = 2a + 4c = 2.$$

Hence there is no such matrix B, so that A is not invertible. We can observe that both the two rows of A are linearly dependent, and also that the two columns of A are linearly dependent. We will later prove that either of these two conditions are equivalent to A *not* being invertible (Corollary 4.93).

Example 4.66. We revisit the reflection, rescaling, rotation and shearing matrices (Example 4.13 onwards) and compute their inverses:

Geometrical description	Matrix A	Inverse matrix A^{-1}
Reflection at x-axis	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
Reflection at y-axis	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$
Rescaling	$\begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}$	$\begin{pmatrix} r^{-1} & 0 \\ 0 & s^{-1} \end{pmatrix}$ (if $r, s \neq 0$;
	if $r = 0$ or $s = 0$ then A is not invertible)	
Rotation	$\begin{pmatrix} \cos r & -\sin r \\ \sin r & \cos r \end{pmatrix}$	$\begin{pmatrix} \cos(-r) & -\sin(-r) \\ \sin(-r) & \cos(-r) \end{pmatrix} =$
Shearing	$\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -r \\ 0 & 1 \end{pmatrix}$ (for any $r \in \mathbf{R}$)

Lemma 4.67. Let A be an invertible matrix. Then there is *precisely* one inverse matrix, i.e., if B and C are two inverses (which means $AB = BA = \mathrm{id}$ and $AC = CA = \mathrm{id}$), then $B = C$. One therefore speaks of *the* inverse (as opposed to *an* inverse), and writes A^{-1} for the inverse.

Proof. Using the associativity of matrix multiplication (marked !), we get the following chain of equalities

$$B = B\mathrm{id} = B(AC) \stackrel{!}{=} (BA)C = \mathrm{id}C = C.$$

Thus $B = C$ as claimed. □

4.9.2 Linear systems associated to invertible matrices

Inverses of matrices are useful to solve linear systems. This is the content of the following theorem:

Theorem 4.68. Let A be an *invertible* $n \times n$-matrix. We consider the linear system

$$Ax = b,$$

where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a vector consisting of n unknowns and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ is a vector. This linear system has a *unique* solution, which is given by

$$x = A^{-1}b,$$

i.e., the product of the *inverse* of A with the vector b.

Remark 4.69. By Observation 4.11, if $A = (a_{ij})$, then the equation $Ax = b$ is a shorthand for the linear system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Proof. We first check that $A^{-1}b$ is indeed a solution to the equation $Ax = b$:

$$A(A^{-1}b) \stackrel{!}{=} (AA^{-1})b = \mathrm{id}_n b = b.$$

At the equation marked ! we have used the associativity of matrix multiplication (where the third matrix is b, which is just a column vector, i.e., an $n \times 1$-matrix).

We now check that $A^{-1}b$ is the *only* solution. Suppose then that some vector y is a solution to the system, i.e., $Ay = b$. We will show $y = A^{-1}b$ by proving

$$z := A^{-1}b - y = 0.$$

Again using the properties of matrix multiplication (Lemma 4.59), we have $Az = A(A^{-1}b - y) = AA^{-1}b - Ay = b - b = 0$. Multiplying this with the matrix A^{-1}, we obtain our claim:

$$z = A^{-1}Az = A^{-1}0 = 0.$$

4.9.3 Structural properties of taking the inverse

Below, it is necessary to have a few properties of the operation "take the inverse of an (invertible) matrix" at our disposal. In the following theorem, all matrices are square matrices of the same size. In the right column, we illustrate what they mean for 1×1-matrices, as a means to remember them. Recall that a 1×1-matrix (a) is invertible if and only if $a \neq 0$, and its inverse $(a)^{-1}$ is the matrix (a^{-1}). (Here, as usual, the *reciprocal* a^{-1} of a non-zero real number a is defined as $a^{-1} := \frac{1}{a}$.)

Theorem 4.70. The following holds:

Statement for general (square) matrices	For 1×1-matrices
id is invertible: $\mathrm{id}^{-1} = \mathrm{id}$	$\frac{1}{1} = 1$
If A is invertible, then A^{-1} is also invertible: $$(A^{-1})^{-1} = A. \qquad (4.71)$$	$\frac{1}{\frac{1}{a}} = a$
If A and B are invertible, then AB is invertible: $$(AB)^{-1} = B^{-1}A^{-1}. \qquad (4.72)$$	$\frac{1}{ab} = \frac{1}{b}\frac{1}{a}$.
If $A_1, \dots, A_k$ are invertible, then their product $A_1 A_2 \dots A_k$ is also invertible: $$(A_1 \dots A_k)^{-1} = A_k^{-1} \dots A_1^{-1} \qquad (4.73).$$	$\frac{1}{a_1 \cdot \dots \cdot a_k} = \frac{1}{a_k} \cdot \dots \cdot \frac{1}{a_1}$

Proof. We prove (4.72) to illustrate the technique. We compute

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} && \text{by associativity} \\ &= AA^{-1} && \text{since } B^{-1} \text{ is inverse of } B \\ &= \mathrm{id} && \text{since } A^{-1} \text{ is inverse of } A. \end{aligned}$$

Using the same arguments, one checks $(B^{-1}A^{-1})(AB) = \mathrm{id}$. Thus, by Definition 4.63, $B^{-1}A^{-1}$ is the inverse of AB. □

Remark 4.74. Among the above formulas, (4.72) is the most noteworthy one: note that the order of A and B has been changed! (Recall from Warning 4.54 that the order of multiplication is important.

Only for 1×1-matrices, i.e., real numbers the order of multiplication is irrelevant, so that $\frac{1}{ab} = \frac{1}{b}\frac{1}{a} = \frac{1}{a}\frac{1}{b}$ etc.)

4.9.4 Invertible matrices and elementary operations

Lemma 4.75. Any elementary matrix (Definition 4.61) is invertible:

$$\left(E^{(1)}_{i,j}\right)^{-1} = E^{(1)}_{i,j}$$
$$\left(E^{(2)}_{i,r}\right)^{-1} = E^{(2)}_{i,r^{-1}}$$
$$\left(E^{(3)}_{i,j,r}\right)^{-1} = E^{(3)}_{i,j,-r}.$$

Proof. To illustrate this, we check this for the last one, where for simplicity of notation we just treat the case of 2×2-matrices. I.e., we prove

$$\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -r \\ 0 & 1 \end{pmatrix}.$$

To do this, we compute the product

$$\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -r \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} & \\ 1 \cdot (-r) + r1 & \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Similarly (or, actually, symmetrically)

$$\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -r \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} & \\ 1 \cdot (-r) + r1 & \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Lemma 4.76. If A is an $m \times n$-matrix and B is obtained from A by means of elementary row operations:

$$A \rightsquigarrow B,$$

then

$$B = UA$$

for an invertible $m \times m$-matrix U. (In particular, if $A \rightsquigarrow \mathrm{id}$, then $\mathrm{id} = UA$.)

Proof. If $A \rightsquigarrow B$ in a single step, this is a combination of Lemma 4.75 and Proposition 4.60: in this case U is the elementary, and in particular invertible, matrix corresponding to the elementary operation that has been performed.

In general, say $A =: A_0 \rightsquigarrow A_1 \rightsquigarrow A_2 \rightsquigarrow \cdots \rightsquigarrow A_n = B$, then $A_1 = U_1 A$, $A_2 = U_2 A_1$ etc., so that

$$A_n = U_n A_{n-1} = U_n U_{n-1} A_{n-2} = \cdots = \underbrace{U_n U_{n-1} \dots U_1}_{=:U} A,$$

where we have used the associativity of matrix multiplication. Being the product of elementary, and in particular invertible matrices, U is then also invertible (Theorem 4.70). □

We finally prove Theorem 2.18. You can check that there is no vicious circle!

Corollary 4.77. Let

$$Ax = b \tag{4.78}$$

be a linear system. Apply any sequence of elementary row operations to A and to b, getting a matrix A' and a vector b'. Then the system

$$A'x = b' \tag{4.79}$$

is equivalent to (4.78), i.e., the solution sets of the two systems are the same.

Proof. By Lemma 4.76, there is an invertible matrix U such that $A' = UA$ and $b' = Ub$. If $Ax = b$, then also

$$A'x = (UA)x = U(Ax) = Ub = b'.$$

Conversely, if $A'x = b'$, then (crucially using that U is invertible)

$$Ax = U^{-1}UAx = U^{-1}A'x = U^{-1}b' = U^{-1}Ub = b.$$

4.9.5 Invertibility criteria

We can now establish a criterion that determines whether a given matrix A is invertible (and that computes the inverse in case it is). This can then be used in practice to apply Theorem 4.68.

Recall that three statements "X", "Y", "Z" are *equivalent* if any of them implies the others. For example the statements (where r is a real number)

- $r + 1 \geqslant 1$
- $r \geqslant 0$
- $r - 4 \geqslant -4$

are equivalent. By contrast, the three statements

- $r + 1 \geqslant 1$
- $r \geqslant 0$
- $r^2 \geqslant 0$

are *not* equivalent, since the third does not imply, say, the second: for $r = -1$, the third statement holds, but the second does not. A convenient way to show that three statements are equivalent is to show "X" $\Rightarrow$ "Y", then "Y" $\Rightarrow$ "Z", and then "Z" $\Rightarrow$ "X". Of course, this also works similarly for more than three statements.

Theorem 4.80. The following conditions on a square matrix $A \in \mathrm{Mat}_{n\times n}$ are equivalent:

(1) A is invertible.

(2) For *any* $b \in \mathbf{R}^n$ (regarded as a column vector with n rows), the equation $Ax = b$ (for $x \in \mathbf{R}^n$ being a column vector consisting of n unknowns $x_1, \dots, x_n$) has *exactly one* solution.

(3) For *any* $b \in \mathbf{R}^n$, the equation $Ax = b$ has *at most one* solution.

(4) The system $Ax = 0$ (0 being the zero row vector consisting of n zeros) has only the trivial solution $x = 0$ (cf. Remark 2.14).

(5) Using the Gaussian algorithm (Method 2.30), A can be transformed to the identity matrix id_n.

(6) A is a product of (appropriate) elementary matrices.

(7) There is a matrix $B \in \mathrm{Mat}_{n\times n}$ such that $AB = \mathrm{id}$.

If these conditions are satisfied, the inverse of A can be computed as follows: write the identity $n \times n$-matrix to the right of A (this gives a $n \times (2n)$-matrix):

$$B := (A \mid \mathrm{id}_n) = \left(\begin{array}{ccc|ccc} a_{11} & \dots & a_{1n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} & 0 & \dots & 1 \end{array}\right).$$

(The bar in the middle is just there for visual purposes, it has no deeper meaning.) Apply Gaussian elimination in order to bring the matrix B to reduced row echelon form, which according to the above gives a matrix of the form

$$(\mathrm{id}_n \mid E).$$

Then $E = A^{-1}$, i.e., E is the inverse of A.

Proof. (1) $\Rightarrow$ (2): This is just the content of Theorem 4.68.

The implications (2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (4) are clear.

(4) $\Rightarrow$ (5): we can bring A into reduced row-echelon form, say, $A \rightsquigarrow R$. We need to show that $R = \mathrm{id}$. If this is not the case, then R contains a zero row (since R is a *square* matrix). Method 2.32 then tells us that the system $Rx = 0$ has (at least) one free parameter, and therefore the system has not only the zero vector as a solution. The original system $Ax = 0$, which by Corollary 4.77 has the same solutions as $Rx = 0$, then also has a non-trivial solution. This is a contradiction to our assumption that R is not the identity matrix.

(5) $\Rightarrow$ (6): by Lemma 4.76, we have $UA = \mathrm{id}$ for U being a product of elementary matrices, say $U = U_1 \dots U_n$. Then, using (4.73), we have

$$A = U^{-1}UA = U^{-1} = U_n^{-1} \dots U_1^{-1},$$

and this is also a product of elementary matrices.

(6) $\Rightarrow$ (7): if $A = U_1 \dots U_n$ for some elementary matrices, then

$$AU_n^{-1} \dots U_1^{-1} = U_1 \dots U_n U_n^{-1} \dots U_1^{-1} = \mathrm{id}.$$

(7) $\Rightarrow$ (1): suppose B is such that $AB = \mathrm{id}$. We observe that then the only vector $x \in \mathbf{R}^n$ such that $Bx = 0$ is the zero vector:

$$x = \mathrm{id}_n x = ABx = A0 = 0.$$

Applying the implication (4) $\Rightarrow$ (7) (which was already proved) to B, we obtain a matrix C such that $BC = \mathrm{id}$. Therefore

$$A = A\mathrm{id}_n = A(BC) = (AB)C = \mathrm{id}C = C.$$

This means that $BA = \mathrm{id}$.

This finishes the proof that all the given statements are equivalent. The statement about the computation of A^{-1} holds since the row operations that bring $A \rightsquigarrow \mathrm{id}$ also bring the augmented matrix $(A \mid \mathrm{id})$ to $(UA \mid U\mathrm{id}) = (\mathrm{id} \mid U)$. □

Example 4.81. We apply this to $A = \begin{pmatrix} 1 & 0 & -1 \\ 3 & 1 & -3 \\ 1 & 2 & -2 \end{pmatrix}$:

$$B = \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 3 & 1 & -3 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array}\right).$$

We subtract the first row 3, resp. 2 times from the other ones, which gives

$$\left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 2 & -1 & -1 & 0 & 1 \end{array}\right).$$

We subtract 2 times the second row from the third:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & -1 & 5 & -2 & 1 \end{array}\right).$$

We bring the matrix into row echelon form by multiplying the last row with -1, which yields

$$\left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & -5 & 2 & -1 \end{array}\right).$$

Finally, to bring it into reduced row-echelon form, we add the third row to the first, which gives

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -4 & 2 & -1 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & -5 & 2 & -1 \end{array}\right).$$

Thus, according to Theorem 4.80, A is indeed invertible, and its inverse is

$$A^{-1} = \begin{pmatrix} -4 & 2 & -1 \\ -3 & 1 & 0 \\ -5 & 2 & -1 \end{pmatrix}.$$

Corollary 4.82. If A is a square matrix such that for some other square matrix B we have $AB = \mathrm{id}$, then we also have $BA = \mathrm{id}$.

Proof. We use the theorem to see that A is invertible, and then

$$B = \mathrm{id} B = A^{-1}AB = A^{-1}$$

And we have seen in (4.71) above that $A^{-1}A = \mathrm{id}$. □

4.10 Change of basis

Let V be a vector space with a basis $v_1, \dots, v_n$. For brevity we write $\underline{v}$ for this basis. Recall from Proposition 3.61 that then any vector $x \in V$ can be written *uniquely* as

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

and we regard the coefficients $\alpha_1, \dots, \alpha_n$ as the coordinates of x with respect to the basis $\underline{v}$. We indicate this notationally by writing $(\alpha_1, \dots, \alpha_n)_{\underline{v}}$.

If we take, instead, another basis $\underline{w}$ consisting of vectors $w_1, \dots, w_n \in V$ then

$$x = \beta_1 w_1 + \cdots + \beta_n w_n,$$

giving *different* coordinates of x with respect to the basis $\underline{w}$. Our goal in this section is to answer the natural question how to pass from the coordinates $(\alpha_1, \dots, \alpha_n)_{\underline{v}}$ to $(\beta_1, \dots, \beta_n)_{\underline{w}}$.

Example 4.83. Consider the identity map $\mathrm{id}_V : V \to V$ (Example 4.7). We fix a basis $\underline{v} = \{v_1, \dots, v_n\}$ of V and determine the matrix of id_V with respect to this basis both in the domain and in the codomain. We have

$$\begin{aligned} \mathrm{id}_V(v_1) &= v_1 = 1 \cdot v_1 + 0v_2 + \cdots + 0v_n \\ &\vdots \\ \mathrm{id}_V(v_n) &= v_1 = 0 \cdot v_1 + \cdots + 0v_{n-1} + 1v_n. \end{aligned}$$

These coefficients in $\mathrm{id}_V(v_i)$ form the i-th column of the matrix, which therefore is equal to

$$\mathrm{M}_{\mathrm{id}_V, \underline{v}, \underline{v}} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix} = \mathrm{id}_n.$$

Example 4.84. We now consider still the identity map id_V, but with a basis $\underline{v} = \{v_1, \dots, v_n\}$ on the domain and another basis $\underline{w} = \{w_1, \dots, w_n\}$ on the codomain. The matrix of id_V with respect to these bases is found by expressing

$$\mathrm{id}_V(v_1) = v_1 = a_{11} w_1 + \cdots + a_{n1} w_n$$

i.e., we express v_1 in coordinates with respect to the basis $\underline{w}$. More generally, for all $i \leqslant n$:

$$\mathrm{id}_V(v_i) = v_1 = a_{1i}w_1 + \cdots + a_{ni}w_n.$$

The matrix of id_V with respect to the bases $\underline{v}$ (domain) and $\underline{w}$ (codomain) is then

$$\mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

We refer to this matrix as the *base change matrix* from $\underline{v}$ to $\underline{w}$.

Example 4.85. Here is a concrete example of the above situation. Let $V = \mathbf{R}^2$, $\underline{v} = \{e_1, e_2\} = \{(1,0),(0,1)\}$ be the standard basis and $\underline{w} = \{w_1, w_2\} = \{(1,1),(1,3)\}$ be another basis. We compute the matrix following the above lines:

$$\mathrm{id}_{\mathbf{R}^2}(1,0) = (1,0) = a_{11}(1,1) + a_{21}(1,3)$$

has the solution $a_{11} = \frac{3}{2}$ and $a_{21} = -\frac{1}{2}$.

$$\mathrm{id}_{\mathbf{R}^2}(0,1) = (0,1) = a_{12}(1,1) + a_{22}(1,3)$$

has the solution $a_{12} = -\frac{1}{2}$, $a_{22} = \frac{1}{2}$, so the matrix reads

$$\mathrm{M}_{\mathrm{id}_{\mathbf{R}^2},\underline{v},\underline{w}} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Thus, for example, the vector $(3,2) = 3e_1 + 2e_2$ can be expressed as

$$\mathrm{M}_{\mathrm{id}_{\mathbf{R}^2},\underline{v},\underline{w}} \begin{pmatrix} 3 \\ 2 \end{pmatrix}_{\underline{v}} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix}_{\underline{v}} = \begin{pmatrix} \frac{7}{2} \\ -\frac{1}{2} \end{pmatrix}_{\underline{w}},$$

i.e.,

$$(3,2) = \frac{7}{2}w_1 - \frac{1}{2}w_2.$$

By Proposition 4.51, the composition of linear maps corresponds to the product of matrices. We apply this to the composition

$$\mathrm{id}_V \circ \mathrm{id}_V = \mathrm{id}_V.$$

We choose the following bases on V, indicated by a subscript:

$$V_{\underline{v}} \stackrel{\mathrm{id}_V}{\to} V_{\underline{w}} \stackrel{\mathrm{id}_V}{\to} V_{\underline{v}}.$$

We consider the associated matrices $\mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}}$ for the first map, and $\mathrm{M}_{\mathrm{id}_V,\underline{w},\underline{v}}$ for the second one. We have

$$\mathrm{M}_{\mathrm{id}_V,\underline{w},\underline{v}} \cdot \mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}} = \mathrm{M}_{\mathrm{id},\underline{v},\underline{v}} = \mathrm{id}_n.$$

In other words,

$$\mathrm{M}_{\mathrm{id}_V,\underline{w},\underline{v}} = (\mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}})^{-1}.$$

The above observations lead to the following.

Method 4.86. Let $f : V \to V$ be a linear map represented by a matrix $E \in \mathrm{Mat}_{n\times n}$ with respect to a fixed basis $\underline{v}$ on the domain and codomain. The matrix of f with respect to another basis $\underline{w}$ (again on the domain and the codomain) is

$$\mathrm{M}_{f,\underline{w},\underline{w}} = AEA^{-1},$$

where A is the matrix describing the change of basis from $\underline{v}$ to $\underline{w}$, i.e., $A = \mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}}$ (cf. Example 4.84).

Proof. Indeed, E is the matrix for $V_{\underline{v}} \stackrel{f}{\to} V_{\underline{v}}$, which we indicate by writing

$$V_{\underline{v}} \underset{E}{\stackrel{f}{\to}} V_{\underline{v}}.$$

The composition

$$V \stackrel{\mathrm{id}_V}{\to} V \stackrel{f}{\to} V \stackrel{\mathrm{id}_V}{\to} V$$

is again f, but as above we now put a different basis on V:

$$V_{\underline{w}} \underset{\mathrm{M}_{\mathrm{id}_V,\underline{w},\underline{v}}}{\stackrel{\mathrm{id}_V}{\to}} V_{\underline{v}} \underset{E}{\stackrel{f}{\to}} V_{\underline{v}} \underset{\mathrm{M}_{\mathrm{id}_V,\underline{v},\underline{w}}}{\stackrel{\mathrm{id}_V}{\to}} V_{\underline{w}}.$$

According to the above, this simplifies to

$$V_{\underline{w}} \underset{A^{-1}}{\stackrel{\mathrm{id}_V}{\to}} V_{\underline{v}} \underset{E}{\stackrel{f}{\to}} V_{\underline{v}} \underset{A}{\stackrel{\mathrm{id}_V}{\to}} V_{\underline{w}}.$$

Example 4.87. Consider the linear map $f : \mathbf{R}^2 \to \mathbf{R}^2$ given in the standard basis $\underline{e} = \{e_1, e_2\}$ by multiplication with the matrix

$E = \begin{pmatrix} 2 & 0 \\ -1 & 3 \end{pmatrix}$. We compute the matrix associated to f with respect to the basis $\underline{w} = \{w_1, w_2\} = \{(1,1),(1,-2)\}$ (on the domain and codomain). According to the above, we need to compute the base change matrix $A = \mathrm{M}_{\mathrm{id},\underline{e},\underline{w}}$ and its inverse $A^{-1} = \mathrm{M}_{\mathrm{id},\underline{w},\underline{e}}$. The matrix A^{-1} can be computed *more easily* than A, because its entries are given by coefficients in the following linear combinations:

$$\mathrm{id}(w_1) = w_1 = (1,1) = 1 \cdot e_1 + 1 \cdot e_2,$$

$$\mathrm{id}(w_2) = w_2 = (1,-2) = 1 \cdot e_1 - 2 \cdot e_2.$$

Thus $A^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. We compute $A = (A^{-1})^{-1}$ using the method described in Theorem 4.80:

$$\begin{aligned} (A^{-1}|\mathrm{id}) &= \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & -2 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & -3 & -1 & 1 \end{array}\right) \\ &\rightsquigarrow \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & \frac{1}{3} & -\frac{1}{3} \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & \frac{1}{3} & -\frac{1}{3} \end{array}\right) \\ &= (\mathrm{id}|(A^{-1})^{-1}) = (\mathrm{id}|A). \end{aligned}$$

This leads to

$$\mathrm{M}_{f,\underline{w},\underline{w}} = AEA^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ -1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix}.$$

4.11 Transposition of matrices

Definition 4.88. If A is an $m \times n$-matrix, then the *transpose* (denoted A^T) is the $n \times m$-matrix obtained by A by reflecting the entries along the main diagonal. More formally, if $A = (a_{ij})$, then

$$A^T := (a_{ji}).$$

Example 4.89. For $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$,

$$A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

The transpose of a column vector is a row vector and vice versa. For example, for $v = \begin{pmatrix} x \\ y \end{pmatrix}$,

$$v^T = \begin{pmatrix} x & y \end{pmatrix}.$$

We have the following basic computation rules involving the transpose.

Lemma 4.90. Let A be an $m \times n$-matrix and $r \in \mathbf{R}$ a real number.

(1) $(A^T)^T = A$, i.e., the transpose of the transpose equals the original matrix,

(2) For another matrix B of the same size as A, $(A+B)^T = A^T + B^T$.

(3) For an $n \times k$-matrix B, the transpose of the matrix product AB is the products of the transposes *in the opposite order*:

$$(AB)^T = B^T A^T. \tag{4.91}$$

(4) If a square matrix A is invertible, then A^T is also invertible with inverse

$$(A^T)^{-1} = (A^{-1})^T. \tag{4.92}$$

Proof. The first two rules are quite immediate to check (and hardly surprising). The first one can also be seen by noting that doing twice the reflection of the entries along the main diagonal gives back the original matrix.

The equation (4.91) is also directly following from the definitions: let $A = (a_{ij})$, $B = (b_{ij})$. Let us write $C = AB = (c_{ij})$. Then $c_{ij} = \sum_{e=1}^n a_{ie} b_{ej}$. Thus $C^T = (c_{ji}) = \sum_{e=1}^n a_{je} b_{ei}$. This equals the (i, j)-entry of $B^T A^T$.

For (4.92), we compute

$$\begin{aligned} A^T (A^{-1})^T &= (A^{-1}A)^T && \text{by (4.91)} \\ &= \mathrm{id}^T \\ &= \mathrm{id}. \end{aligned}$$

Similarly,

$$\begin{aligned} (A^{-1})^T A^T &= (AA^{-1})^T && \text{by (4.91)} \\ &= \mathrm{id}^T \\ &= \mathrm{id}. \end{aligned}$$

Thus the product of A^T and $(A^{-1})^T$ (in the two possible orders) equals id, so they are inverse to each other. □

The usage of transposes helps us prove another set of equivalent characterizations:

Corollary 4.93. Let $A \in \mathrm{Mat}_{n\times n}$ be a square matrix. Then the following are equivalent:

(1) A is invertible.

(2) The n columns of A are linearly independent.

(3) The rank of A is n.

(4) The n rows of A are linearly independent.

Proof. (1) ⇔ (2): According to Theorem 4.80, A is invertible precisely if the only solution to the system $Ax = 0$ is the zero vector $x = 0$. Recalling that for $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ we have

$$Ax = x_1c_1 + \cdots + x_nc_n,$$

where $A = (c_1 \ \dots \ c_n)$ are the columns of A, we see that the above condition is equivalent to the columns being linearly independent.

(2) ⇔ (3): The rank is, by definition, the dimension of the column space, i.e., the subspace of $\mathbf{R}^n$ generated by the columns $c_1, \dots, c_n$. In order to show that these vectors span $\mathbf{R}^n$, let $b \in \mathbf{R}^n$. By the invertibility of A, we know that the system $Ax = b$ has a (unique) solution x. Therefore $Ax = \sum_{k=1}^n x_ic_i = b$.

(1) ⇔ (4): A is invertible if and only if the transpose A^T is invertible. Now use that the rows of A^T are the columns of A, and apply the (already proved) equivalence (1) ⇔ (2). □

4.12 Exercises

Exercise 4.1. Determine which 2×2-matrix A is such that the function

$$f : \mathbf{R}^2 \to \mathbf{R}^2, v \mapsto Av$$

are the following:

- $f(v)$ is the point v reflected along the y-axis,

- $f(v)$ is the same point as v,
- $f(v)$ is the origin $(0,0)$,
- $f(v)$ is the point v reflected along the line $\{(x,x) \mid x \in \mathbf{R}\}$ (i.e., the "southwest-northeast diagonal"),
- $f(v)$ is the point v rotated counterclockwise, resp. clockwise by $60°$?

Exercise 4.2. Determine the matrix A such that $Av = \begin{pmatrix} -y \\ x \end{pmatrix}$. Describe the behaviour of the function $v \mapsto Av$ geometrically.

Exercise 4.3. Write down the matrix A such that the function $f : \mathbf{R}^4 \to \mathbf{R}^3, v \mapsto Av$ satisfies

$$f((1,0,0,0)) = (1,2,3), f((0,1,0,0)) = (0,0,7),$$
$$f((0,0,1,0)) = (0,0,0), f((0,0,0,1)) = (13,0,-1).$$

Determine $\ker f$ and $\operatorname{im} f$ (i.e., determine a basis and their dimension).

Exercise 4.4. Compute the rank of

$$A = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 1 & 1 & 4 & 2 \end{pmatrix}.$$

Exercise 4.5. Consider the linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ described in Example 4.42. Determine the matrix of f with respect to the standard basis e_1, e_2, e_3 (both in the "source" $\mathbf{R}^3$, and also in the "target" $\mathbf{R}^3$).

Exercise 4.6. For $\lambda \in \mathbf{R}$ consider the subspace of $\mathbf{R}^3$ defined as

$$W_\lambda = L((1, 1+\lambda, -1), (2, \lambda-2, \lambda+2)).$$

For each $\lambda \in \mathbf{R}$, find a basis and the dimension of W_λ.

Exercise 4.7. Determine the rank of

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 1+\alpha \\ \alpha & 1 & 2 \end{pmatrix}$$

for each $\alpha \in \mathbf{R}$.

Exercise 4.8. For any $\lambda \in \mathbf{R}$ solve the system (in the unknowns x_1, x_2, x_3)

$$\begin{aligned} \lambda x_1 &= 0 \\ \lambda x_2 + (1+\lambda)x_3 &= 1 \\ \lambda x_1 + x_2 + 2x_3 &= 3. \end{aligned}$$

Exercise 4.9. (Solution at p. 244) Consider the linear map

$$f : \mathbf{R}^2 \to \mathbf{R}^3, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_2 \\ 3x_1 + 5x_2 \end{pmatrix}.$$

(1) Determine $\ker f$.

(2) Does the vector $\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$ lie in the image of F?

Exercise 4.10. (Solution at p. 245) Consider the linear map

$$f : \mathbf{R}^4 \to \mathbf{R}^3$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \mapsto \begin{pmatrix} 2 & -1 & 1 & 1 \\ 0 & 5 & -3 & -5 \\ 3 & -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 + x_3 + x_4 \\ 5x_2 - 3x_3 - 5x_4 \\ 3x_1 - 4x_2 + 3x_3 + 4x_4 \end{pmatrix}.$$

(1) Determine $\ker f$.

(2) Determine $f^{-1}(\begin{pmatrix} 1 \\ -3 \\ -3 \end{pmatrix})$, i.e., find all the vectors $v \in \mathbf{R}^4$ such that $f(v) = \begin{pmatrix} 1 \\ -3 \\ -3 \end{pmatrix}$. Is this subset of $\mathbf{R}^4$ a subspace?

Exercise 4.11. (Solution at p. 245) Consider the matrix

$$A_t = \begin{pmatrix} 1 & 3 & -1 & 2 \\ 1 & 5 & 1 & 1 \\ 2 & 4 & t & 5 \end{pmatrix}.$$

Here $t \in \mathbf{R}$ is an arbitrary real number.

(1) Determine the rank of A_t.

(2) Set $t = -4$, $u = \begin{pmatrix} 1 \\ \alpha \\ 0 \end{pmatrix}$. Find $\alpha \in \mathbf{R}$ such that

$$A_{-4} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = u$$

has solutions.

(3) Set again $t = -4$, $v = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}$. Determine the solutions of the linear system

$$A_{-4} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = v.$$

(4) Is there any $t \in \mathbf{R}$ such that the homogeneous system

$$A_t \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has only the trivial solution $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$?

Exercise 4.12. Let $f : V \to W$ be a linear map. For a subspace $U \subset V$ we define the *image* of U to be

$$f(U) := \{f(v) \mid v \in U\}.$$

(For example, for $U = V$, this gives back the image of f as defined in Definition 4.20).

(1) Arguing as in Proposition 4.23, prove that $f(U)$ is a subspace of W.

(2) Prove that $\dim f(U) \leqslant \dim U$.

Exercise 4.13. Let $f : V \to W$ be a linear map. For a subspace $U \subset W$, we define the *preimage* of U to be

$$f^{-1}(U) := \{v \in V \mid f(v) \in U\}.$$

(For example, if $U = \{0_W\}$ is the subspace consisting only of the zero vector of W, this gives back the kernel: $\ker f = f^{-1}(\{0_W\})$.)

Arguing as in Proposition 4.23, prove that $f^{-1}(U)$ is a subspace of V.

Exercise 4.14. (Solution at p. 246) Consider the linear map $f : \mathbf{R}^4 \to \mathbf{R}^3$ given by multiplication with the matrix

$$A = \begin{pmatrix} 2 & -1 & -\frac{5}{2} & 1 \\ -1 & 0 & 1 & -\frac{1}{2} \\ 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 0 \end{pmatrix}.$$

Determine $\ker f$, $\operatorname{im} f$ and $\ker f \cap \operatorname{im} f$.

Exercise 4.15. Consider the linear map

$$f : \mathbf{R}^3 \to \mathbf{R}^2$$

given by

$$f(x, y, z) = (2x - z, x + y + z).$$

(1) Determine the matrix of f with respect to the standard basis in $\mathbf{R}^3$ and the standard basis $\mathbf{R}^2$.

(2) Determine $\ker f$ and $\operatorname{im} f$.

(3) Determine the preimage $f^{-1}((0, 1))$. Write down the linear system whose solution set is this preimage. Is it a subspace of $\mathbf{R}^3$?

(4) Show that the vectors $v_1 = (0, 1, 2)$, $v_2 = (0, -1, 1)$ and $v_3 = (1, 1, 1)$ are a basis of $\mathbf{R}^3$. Determine the matrix of f with respect to this basis of $\mathbf{R}^3$ and the standard basis in the codomain $\mathbf{R}^2$.

Exercise 4.16. (Solution at p. 248) Consider the linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ whose matrix with respect to the standard basis (of both the domain and the codomain $\mathbf{R}^3$) is

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix}.$$

(1) Let $v_1 = (1, 1, 0)$. Compute $v_2 = f(v_1)$ and $v_3 = f(v_2)$. Show that v_1, v_2, v_3 form a basis of $\mathbf{R}^3$.

(2) Consider $v_4 = f(v_3)$ and determine a_1, a_2, a_3 such that

$$v_4 = a_1 v_1 + a_2 v_2 + a_3 v_3.$$

(3) Determine the matrix of f with respect to the basis v_1, v_2, v_3 (both of the domain and of the codomain $\mathbf{R}^3$).

Exercise 4.17. (Solution at p. 249) Consider the linear map

$$f : \mathbf{R}^4 \to \mathbf{R}^3$$
$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \mapsto \begin{pmatrix} -x+z \\ -y+t \\ x-y \end{pmatrix}$$

- Write the matrix associated to f with respect to the standard basis of the domain $\mathbf{R}^4$ and the standard basis of the codomain $\mathbf{R}^3$.
- Determine $\ker f$ and $\operatorname{im} f$.

Exercise 4.18. (Solution at p. 250) Consider the following functions:

$$f_1 : \mathbf{R}^2 \to \mathbf{R}^3, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_2 + x_1 \\ 3x_1 \\ 2x_2 \end{pmatrix}$$

$$f_2 : \mathbf{R}^3 \to \mathbf{R}^2, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_3 + x_1 \\ 3x_2 + 4x_1 + 1 \end{pmatrix}$$

$$f_3 : \mathbf{R}^3 \to \mathbf{R}^2, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_3^2 \\ x_2 + x_1 \end{pmatrix}$$

$$f_4 : \mathbf{R}^3 \to \mathbf{R}^3, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$f_5 : \mathbf{R}^3 \to \mathbf{R}^3, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + x_2 \\ x_2 + x_3 \\ x_3 + x_1 \end{pmatrix}$$

(1) Verify which ones are linear.

(2) Determine all the images.

(3) Determine if the described linear functions are injective or/and surjective.

Exercise 4.19. Consider the vectors in $\mathbf{R}^4$

$$\begin{aligned} v_1 &= (1, 2, -3, -1) \\ v_2 &= (3, 4, 4, 1) \\ v_3 &= (1, 0, 10, 3). \end{aligned}$$

(1) Are v_1, v_2, v_3 linearly independent? What is the dimension of the subspace U of $\mathbf{R}^4$ that these vectors span?

(2) Find a basis of $\mathbf{R}^4$ that contains at least 2 of these three vectors.

(3) Define a map $f : \mathbf{R}^4 \to \mathbf{R}^2$ that satisfies $f(U) = \mathbf{R}^2$. Can you define a map $g : \mathbf{R}^4 \to \mathbf{R}^3$ that satisfies $g(U) = \mathbf{R}^3$?

Example 4.94. Decide whether

$$A = \begin{pmatrix} 1 & 3 & -1 \\ 2 & 1 & 5 \\ 1 & -7 & 13 \end{pmatrix}$$

is invertible. If A is invertible, determine its inverse.

Exercise 4.20. Decide whether $AB = BA$ holds for the following matrices:

(1) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$

(2) $A = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$, $B = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$

(3) $A = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} y & 1 \\ 0 & y \end{pmatrix}$ for two fixed real numbers x and y

(4) $A = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}$ an arbitrary 2×2-matrix

(5) A an arbitrary matrix, $B = AA$ (the product of A with itself, this is also denoted A^2)

Exercise 4.21. Determine in each case whether there is a matrix A satisfying the following condition. If so, is there a unique such matrix or can there be several matrices satisfying the condition? Describe your findings geometrically.

- $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$
- $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $A\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$
- $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $A\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

Exercise 4.22. Find two 2×2-matrices A, B such that

$$AB = 0$$

(the zero matrix), but $A \neq 0$ and $B \neq 0$. (Hint: start with $A = B = 0$, and then change very few entries.)

Exercise 4.23. A square matrix A is called *symmetric* if

$$A = A^T.$$

(1) Determine $s, t \in \mathbf{R}$ such that the matrix $\begin{pmatrix} 1 & s \\ -2 & t \end{pmatrix}$ is symmetric.
(2) Let A be any square matrix. Prove that $A + A^T$ is always symmetric. (Hint: Use Lemma 4.90).

Exercise 4.24. The *trace* of a square matrix $A = (a_{ij})$ is defined to be the sum of the entries on the main diagonal:

$$\operatorname{tr}(A) := a_{11} + a_{22} + \cdots + a_{nn}.$$

Prove the following statements (if you get stuck with the notation, assume first that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2-matrix, then $\operatorname{tr}(A) = a + d$):
(1) $\operatorname{tr}(A) = \operatorname{tr}(A^T)$,
(2) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ (for another square matrix B of the same size). This is noteworthy since $AB \neq BA$ in general!

(3) $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ (for another square matrix B of the same size), $\mathrm{tr}(rA) = r\mathrm{tr}(A)$ (for $r \in \mathbf{R}$).

(4) (optional, slightly more challenging) Prove there is no matrix B such that $AB - BA = \mathrm{id}$.

Exercise 4.25. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an arbitrary 2×2-matrix and $r \in \mathbf{R}$. Recall that the scalar multiple $rA = \begin{pmatrix} ra & rb \\ rc & rd \end{pmatrix}$. Find a 2×2-matrix R such that the matrix product RA equals the scalar multiple:

$$RA = rA.$$

Exercise 4.26. (1) Let

$$A = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$$

be a so-called *upper triangular matrix* (of size 3×3). Compute A^2 and prove that $A^3 = 0$.

(2) Make a (sensible) similar statement for $n \times n$-matrices (cf. Proposition 5.17 for the definition of upper triangular matrices in general).

Exercise 4.27. (Solution at p. 251) Consider the identity map

$$\mathrm{id} : \mathbf{R}^2 \to \mathbf{R}^2, (x, y) \mapsto (x, y).$$

Consider the standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$ of the domain, and the basis comprised of $v_1 = (1, -3)$ and $v_2 = (2, 1)$ on the codomain.

- Compute the base change matrix of id with respect to these bases.
- Use it to compute the coordinates of $(2, -5)$ in terms of the basis v_1, v_2.

Exercise 4.28. (Solution at p. 252) Consider the identity map

$$\mathrm{id} : \mathbf{R}^3 \to \mathbf{R}^3$$

and the basis $v_1 = (1, 0, -1)$, $v_2 = (2, 1, 1)$, $v_3 = (-1, -1, 7)$ on the domain and the standard basis on the codomain. Compute the base change matrix with respect to these bases.

Exercise 4.29. Find the base change matrix from the standard basis e_1, e_2, e_3 in $\mathbf{R}^3$ to the basis $v_1 = (1, 1, 2)$, $v_2 = (1, 1, 3)$, $v_3 = (7, -1, 0)$.

Exercise 4.30. (Solution at p. 252) Consider the linear map

$$f : \mathbf{R}^2 \to \mathbf{R}^3, v \mapsto Av,$$

where $A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ -3 & 1 \end{pmatrix}$. Compute the matrix B of f with respect to the basis $\underline{v} = \{v_1 = (1, -1), v_2 = (3, -1)\}$ in $\mathbf{R}^2$, and the basis $\underline{t} = \{t_1 = (1, 0, 1), t_2 = (2, 1, 1), t_3 = (-1, -1, -1)\}$ in $\mathbf{R}^3$.

Hint: We may consider the following diagram:

$$\mathbf{R}^2_{\underline{v}} \xrightarrow[H]{\text{id}} \mathbf{R}^2_{\underline{e}} \xrightarrow[A]{f} \mathbf{R}^3_{\underline{e}} \xrightarrow[K]{\text{id}} \mathbf{R}^3_{\underline{t}}.$$

Here the subscripts at $\mathbf{R}^2$ indicate which basis we consider. The matrices H and K are the base change matrices from the basis $\underline{v}$ to the standard basis $\underline{e}$, resp. from the standard basis $\underline{e}$ to the basis $\underline{t}$. Then

$$B = K \cdot A \cdot H.$$

Exercise 4.31. Consider the linear map

$$f : \mathbf{R}^3 \to \mathbf{R}^3$$

which in the standard basis (on both the domain and the codomain) is given by

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Compute the matrix of f with respect to the basis

$$v_1 = (0, -1, 1), v_2 = (0, 1, 0), v_3 = (-1, 0, 1).$$

(on the domain and the codomain).

Exercise 4.32. (Solution at p. 253) Consider the linear map

$$f : \mathbf{R}^2 \to \mathbf{R}^2,$$

which is given by the matrix $A = \begin{pmatrix} 6 & -1 \\ 2 & 3 \end{pmatrix}$ with respect to the standard basis in the domain and the codomain.

Find its matrix with respect to the basis $v_1 = (1, 1)$, $v_2 = (1, 2)$ both in the domain and the codomain.

Exercise 4.33. (Solution at p. 253) Let $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. Determine the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ such that

$$Ax = x.$$

Exercise 4.34. (Solution at p. 254) Find, if possible the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} x = 5x.$$

Exercise 4.35. (Solution at p. 255) Consider the matrix $A = \begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix}$ which represents $f : \mathbf{R}^2 \to \mathbf{R}^2$ with respect to the standard basis. Find the matrix of f with respect to the basis

$$\underline{v} = \{v_1 = (2, 1), v_2 = (0, 1)\}.$$

Exercise 4.36. For A and f as in Exercise 4.35, consider now the basis

$$\underline{v} = \{v_1 = (1, 2), v_2 = (0, 1)\}.$$

Compute the matrix of f with respect to that basis.

Exercise 4.37. (Solution at p. 255) Let $f : \mathbf{R}^3 \to \mathbf{R}^3$ be the map whose matrix with respect to the standard basis is

$$\begin{pmatrix} -1 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & -2 \end{pmatrix}.$$

Compute the matrix with respect to the basis

$$\underline{v} = \{v_1 = (0,0,1), v_2 = (2,6,-1), v_3 = (1,0,1)\}.$$

Exercise 4.38. (Solution at p. 255) Let $f : \mathbf{R}^3 \to \mathbf{R}^3$ be the linear map whose matrix with respect to the standard basis is

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

(1) Find a basis of the solution space L of the linear system

$$A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 3 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

(As a forecast to terminology introduced later, this solution space is the so-called *eigenspace* of A for the *eigenvalue* 3, cf. Definition and Lemma 6.11.)

(2) Complete the basis of L (which is a subspace of $\mathbf{R}^3$) to a basis of $\mathbf{R}^3$, and compute the matrix of f with respect to this basis.

Exercise 4.39. (Solution at p. 256) Consider the linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ whose kernel is $L((1,0,1))$ and such that

$$f(0,3,-1) = (0,3,-1), f(0,0,1) = (0,0,2).$$

Compute its matrix with respect to the standard basis.

Exercise 4.40. Consider the linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ such that

$$f(1,1,1) = 3 \cdot (1,1,1), f(2,0,1) = (-4,0,-2), f(0,1,3) = (0,2,6).$$

(1) Show that the vectors

$$\underline{v} = \{(1,1,1), (2,0,1), (0,1,3)\}$$

form a basis of $\mathbf{R}^3$. (Note that for each of these three vectors, one has $f(v_i) = \lambda_i v$, with $\lambda_1 = 3$ etc. Therefore, the basis is an example of a so-called *eigenbasis*, cf. Definition 6.17.)

(2) Compute the matrix of f with respect to that basis.

(3) Compute the matrix of f with respect to the standard basis.

The following two exercises are all concerned with linear systems of the form

$$Ax = \lambda x,$$

where A is a certain square matrix, x is a vector and $\lambda \in \mathbf{R}$ a real number. We will study these systems systematically in §6.

Exercise 4.41. (Solution at p. 258) Find the solutions of the linear system

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 5 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Exercise 4.42. Solve the system

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = 1 \cdot \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

Exercise 4.43. (Solution at p. 258) Consider the vectors $v_1 = (1, 0, -1)$, $v_2 = (1, 1, 0)$, $v_3 = (1, 0, -2) \in \mathbf{R}^3$.

Let $f\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the linear map such that $f(v_1) = (3, 0, -5)$, $f(v_2) = (2, 2, 0)$ (in the terminology of Definition 6.1, $v_2$is an eigenvector with eigenvalue 2) and $f(v_3) = (5, 2, -5)$.

(1) Determine the matrix A of f with respect to the basis v_1, v_2, v_3 (both in the domain and the codomain).

(2) Determine the matrix B of f with respect to the standard basis of $\mathbf{R}^3$.

(3) Compute a basis of the kernel and of the image of f.

(4) Determine for which value of t is the vector $w = (3, t, -5)$ in the image of f. For such t, compute the preimage $f^{-1}(w)$.

Exercise 4.44. (Solution at p. 260) Let $f\colon \mathbf{R}^3 \to \mathbf{R}^4$ be the following linear map:

$$f(x, y, z) = (x - y + 2z,\ -2x + 3y - z,\ y + 3z,\ -x + 3y + tz)$$

(1) Determine the dimension of the image of f (depending on the parameter $t \in \mathbf{R}$).

(2) Are there values of t for which f is surjective? If so, what are these values of t? Are there values of t for which f is injective? If so, what are they?

(3) For the value of t for which the rank of f is 2, compute a basis of $\ker f$ and of $\Im f$.

(4) Are there values of t for which the vector $w = (1, 1, 0, 1)$ belongs to the image of f?

(5) We now put $t = 0$. Is there a linear map $g \colon \mathbf{R}^4 \to \mathbf{R}^3$ such that the composite $g \circ f \colon \mathbf{R}^3 \to \mathbf{R}^3$ is the identity?

Chapter 5

Determinants

The determinant of a square matrix A is a number that encodes important information about A. For example, A is invertible (Definition 4.63) if and only if the determinant of A is nonzero (Theorem 5.13).

5.1 Determinants of 2×2-matrices

We begin with the definition of determinants of 2×2-matrices.

Definition 5.1. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a 2×2-matrix. The *determinant* of A is defined as

$$\det A := ad - bc.$$

Example 5.2.

- $\det \begin{pmatrix} 4 & 7 \\ 2 & -1 \end{pmatrix} = 4 \cdot (-1) - 7 \cdot 2 = -18.$
- $\det \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} = 1 \cdot 1 - r \cdot 0 = 1.$ In particular, $\det \mathrm{id}_2 = 1$.
- Consider a matrix $A = \begin{pmatrix} a & b \\ ra & rb \end{pmatrix}$ whose second column is a multiple of the first (so that the columns are linearly dependent). Then

$$\det \begin{pmatrix} a & b \\ ra & rb \end{pmatrix} = arb - bra = 0.$$

According to Corollary 4.93, A is *not* invertible. This is an example of the fact alluded to above (cf. Theorem 5.13).

Determinants carry the following geometric meaning. Recall that the *absolute value* of a real number r is defined as

$$|r| := \begin{cases} r & r \geqslant 0 \\ -r & r < 0. \end{cases}$$

For example, $|4| = 4$ and $|-5| = 5$.

Lemma 5.3. Let

$$A = (v \ v') = \begin{pmatrix} x & x' \\ y & y' \end{pmatrix}$$

be a 2×2-matrix, where v and $v' \in \mathbf{R}^2$ are the two columns of A. Then

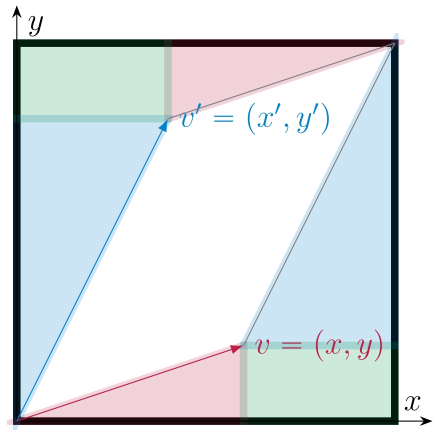
$$|\det(A)| = \operatorname{area}(v_1, v_2),$$

where the right hand side denotes the area of the parallelogram spanned by the two vectors v_1, v_2.

Proof. We illustrate this geometrically in the case where all entries of A are positive and the vectors v and v' lie as depicted, i.e., the angle from v to v' goes, informally speaking, counterclockwise. The area of the black rectangle is $(x + x')(y + y')$. The area of the two triangles whose long side is v (resp. parallel to it), is $\frac{xy}{2}$, so the area of these two triangles together is xy. Likewise the total area of the triangles (parallel to) v' is $x'y'$. Finally, the area of the rectangle at the bottom right, resp. top left corner of the large rectangle is $x'y$. Therefore, the area of the parallelogram is

$$\begin{aligned}(x + x')(y + y') - xy - x'y' - 2x'y &= xy + x'y + xy' + x'y' - xy - x'y' - 2x'y \\ &= xy' - x'y \\ &= \det A.\end{aligned}$$

∎



□

Lemma 5.3 does not give any information about the sign of the determinant. Regarding that, we observe the following:

Lemma 5.4. Let $A = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$ and let

$$A' := \begin{pmatrix} x_2 & x_1 \\ y_2 & y_1 \end{pmatrix}$$

$$A'' := \begin{pmatrix} y_1 & y_2 \\ x_1 & x_2 \end{pmatrix}$$

be the matrices obtained from A by swapping the two columns, resp. the two rows. Then

$$\det A'' = \det A' = -\det A.$$

In other words, swapping two rows or two columns will change the sign of the determinant.

Proof. This is directly clear from the definition. For example,

$$\det A' = x_2 y_1 - y_2 x_1 = -(x_1 y_2 - x_2 y_1) = -\det A.$$

Thus, the determinant (as opposed to only its absolute value) records the area of the parallelogram spanned by the vectors and also the orientation.

5.2 Determinants of larger matrices

There are various (equivalent) approaches to defining determinants of larger matrices. The following one is satisfactory from both a conceptual and a practical standpoint.

Theorem 5.5. There is a *unique* function, called the *determinant*,

$$\det : \mathrm{Mat}_{n\times n} \to \mathbf{R}$$

with the following properties (throughout $A \in \mathrm{Mat}_{n\times n}$):
(1) $\det(\mathrm{id}_n) = 1$,

(2) If A' results from A by interchanging two rows, then

$$\det(A') = -\det(A). \tag{5.6}$$

(3) Let us write a matrix as $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, i.e., $v_i \in \mathbf{R}^n$ is the i-th row of the matrix. Then for any $w \in \mathbf{R}^n$ and any $r \in \mathbf{R}$:

$$\det(\begin{pmatrix} v_1 \\ \vdots \\ rv_i + w \\ \vdots \\ v_n \end{pmatrix}) = r\det(\begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix}) + \det(\begin{pmatrix} v_1 \\ \vdots \\ w \\ \vdots \\ v_n \end{pmatrix}).$$

Remark 5.7. The above operations are somewhat like elementary operations (Definition 2.29): if we take $w = 0$ above, then the formula says that multiplying any one row by r (which may be zero, unlike in Definition 2.29), then the determinant also gets multiplied by r. In particular, if A has a zero row, then

$$\det A = 0. \tag{5.8}$$

Remark 5.9. We also have

$$\det A = 0$$

whenever two rows of A are equal: indeed, the matrix A' obtained by interchanging these rows is equal to A, i.e., $A' = A$, so that $\det A = \det A'$. However, according to (5.6), we also have $\det A' = -\det A$. Taking this together, we have

$$\det A = -\det A$$

and this is only possible if $\det A = 0$.

Remark 5.10. The preceding remark also implies that for $i \neq j$ and $r \in \mathbf{R}$

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_i + r v_j \\ \vdots \\ v_n \end{pmatrix} = \det \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} + r \det \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} \quad \text{where } v_j \text{ is in the } i\text{-th row!}$$

$$= \det \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} + r \begin{pmatrix} \vdots \\ v_j \\ \vdots \\ v_j \\ \vdots \end{pmatrix}$$

$$= \det \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} \quad \text{by the above remark.}$$

∎

In other words, adding an arbitrary multiple of some row to another row does not affect the determinant.

In order to get a feeling for this theorem, let us apply it to a concrete matrix, say

$$A = \begin{pmatrix} -2 & 1 & 8 \\ 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix}.$$

Taking the theorem for granted, we will compute $\det A$ by stepwise applying the above rules and keeping track of how the determinant changes.

$$\underbrace{\begin{pmatrix} -2 & 1 & 8 \\ 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix}}_{=A} \rightsquigarrow \underbrace{\begin{pmatrix} 0 & 7 & 18 \\ 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix}}_{=A_1} \rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 6 \\ 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix}}_{=A_2} \rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 6 \\ 1 & 3 & 5 \\ 0 & 0 & -8 \end{pmatrix}}_{=A_3}$$
$$\rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 6 \\ 1 & 3 & 5 \\ 0 & 0 & 1 \end{pmatrix}}_{=A_4} \rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 6 \\ 1 & 3 & 5 \\ 0 & 0 & 1 \end{pmatrix}}_{=A_5} \rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=A_6}$$
$$\rightsquigarrow \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=A_7} \rightsquigarrow \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=A_8=\mathrm{id}_3}$$

From A to A_1 to A_2 to A_3, we have added appropriate multiples of some row to another one, so that

$$\det A = \det A_1 = \det A_2 = \det A_3.$$

We obtain A_4 from A_3 by multiplying the last row with $-\frac{1}{8}$, so that $\det A_4 = -\frac{1}{8}\det A_3$. From A_4 to A_5 to A_6 to A_7, we again added appropriate multiples to some other rows, so that

$$\det A_4 = \det A_5 = \det A_6 = \det A_7.$$

Finally, A_8 is obtained from A_7 by swapping the first two rows, so that

$$1 = \det A_8 = -\det A_7.$$

Taking this all together we see that

$$\det A = \det A_3 = -8\det A_4 = -8\det A_7 = +8\det A_8 = 8.$$

This shows that the above abstract description of the determinant can be used to compute determinants in practice.

Proof. (of Theorem 5.5) We only sketch the proof idea: one basically proceeds, for a general square matrix, similarly to the computation above: one uses Gaussian elimination, i.e., elementary row operations to bring a given square matrix A into reduced row-echelon form, say $A \rightsquigarrow A'$. The properties in Theorem 5.5 then imply how to compute $\det A$ in terms of $\det A'$. If the resulting matrix A' has a zero row, then $\det A' = 0$. If it has no zero row, then $A' = \mathrm{id}$, and $\det A' = 1$. $\square$

5.2.1 Small matrices

For practical purposes, it is useful to have an explicit formula at hand for small matrices:

(1) For a 1×1-matrix A, i.e., $A = (a)$, we have

$$\det A = a.$$

(2) The determinant of 2×2-matrices defined in Definition 5.1 satisfies the properties listed in Theorem 5.5.

- $\det \mathrm{id}_2 = \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \cdot 1 - 0 \cdot 0 = 1.$
- Swapping two rows yields a sign change in the determinant (Lemma 5.4).
- $$\begin{aligned} \det \begin{pmatrix} a & b \\ c + re & d + rf \end{pmatrix} &= a(d + rf) - b(c + re) \\ &= ad - bc + r(af - be) \\ &= \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} + r \det \begin{pmatrix} a & b \\ e & f \end{pmatrix}. \end{aligned}$$

Thus, the definition of det for general matrices agrees with the one in Definition 5.1.

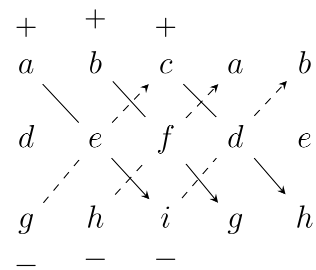
(3) For a 3×3-matrix one can show that the determinant is given by the so-called *Sarrus' rule*:

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - ceg - bdi - afh. \quad (5.11)$$

A way to remember this formula is to write

$$\begin{pmatrix} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{pmatrix}$$

and take products of entries along the top-left-to-bottom-right diagonals with a positive sign, and the top-right-to-bottom-left diagonals with a negative sign:



One can prove by direct computation, that the function defined in (5.11) satisfies the conditions in Theorem 5.5.

(4) Sarrus' rule does not apply to larger matrices. Instead, for matrices of size 4×4, one can prove that $\det A$ is the sum of 24 expressions, each of which is a product of 4 entries of A.

5.3 Invertibility and determinants

We can use the properties of the determinant in Theorem 5.5 (and its proof) to obtain a useful criterion to decide when a matrix is invertible. Determinants can also be used to compute the inverse of an invertible matrix, however this is only of theoretical significance due to the complexity of the ensuing (iterative) algorithm.

Definition 5.12. Let $A \in \mathrm{Mat}_{n\times n}$. For $1 \leqslant i, j \leqslant n$, denote by A_{ij} that is obtained from A by deleting the i-th row and the j-th column. The number

$$c_{ij} := (-1)^{i+j} \det A_{ij}$$

is called the (i, j)-*cofactor* of A.

The *adjugate* of A is the $n \times n$-matrix defined as

$$\mathrm{adj}(A) = (c_{ij}(A))^T = (c_{ji}(A)).$$

Theorem 5.13. An $n \times n$-matrix A is invertible if and only if

$$\det A \neq 0.$$

If this is the case, then the inverse can be computed as

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A).$$

Proof. We revisit the proof of Theorem 5.5: say $A \rightsquigarrow A'$, a reduced row echelon matrix, by means of elementary operations (Gaussian elimination). In this process, one does not multiply any row by zero, so that $\det A = 0$ if and only if $\det A' = 0$. We also know that $A' = UA$, where U is an invertible matrix (namely, a product of elementary matrices). Moreover, A' is invertible if and only if A is invertible (since U is invertible). We therefore have

$$A \text{ is invertible} \Leftrightarrow A' \text{ is invertible} \overset{?}{\Leftrightarrow} \det A' \neq 0 \Leftrightarrow \det A \neq 0$$

and it remains to show the middle equivalence.

The matrix A' is in reduced row echelon form. Thus, either $A' = \mathrm{id}$ or A' contains a zero row. In the first event, A' is invertible, and $\det A' = 1 \neq 0$. In the second event, A' is not invertible (by Corollary 4.93) and $\det A' = 0$ as was noted around (5.8).

We skip the proof of the adjugate formula, cf. [Nic95, Theorem 3.2.4]. □

Example 5.14. For a 2×2-matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the adjugate matrix is

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Therefore, the inverse can be computed as

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

5.4 Further properties of determinants

Proposition 5.15. (*Product formula*) For two $n \times n$-matrices A, B, we have the following formula:

$$\det(AB) = \det A \cdot \det B.$$

I.e., the determinant of a product (of square matrices of the same size) is the product of the two individual determinants.

In particular, this shows

$$\det(AB) = \det(BA),$$

even though $AB \neq BA$!

Proof. We don't include a full proof, but only observe that one checks this by direct computation when A is an elementary matrix. This implies the formula if A is invertible, since then A is a product of elementary matrices. If A is not invertible, then one shows that AB is also not invertible (for any B), and therefore both $\det A = 0$ and $\det(AB) = 0$, so the formula holds in this case too. See, e.g., [Nic95, Theorem 3.2.1] for a proof. □

Remark 5.16. The determinant is therefore multiplicative, but it is *not* additive: one has

$$\det(A + B) \neq \det A + \det B,$$

e.g.

$$\det \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 4 \neq 1 + 1 = \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Proposition 5.17. Let A be an *upper triangular matrix* or a *lower triangular matrix*, i.e., of the form

$$A = \begin{pmatrix} a_{11} & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & a_{nn} \end{pmatrix},$$

resp.

$$A = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ * & \dots & * & a_{nn} \end{pmatrix}.$$

Here * stands for an arbitrary entry. Then

$$\det A = a_{11} \cdot \dots \cdot a_{nn}.$$

Proof. If one of the entries on the main diagonal, i.e., $a_{11}, \dots, a_{nn}$ is zero, then the columns of A are linearly dependent, so that A is not invertible and $\det A = 0$. If instead all $a_{ii} \neq 0$, we can divide the i-th row by a_{ii}, and assume the entries on the main diagonal are all 1. Then, adding appropriate multiples of the rows to the rows above (resp. below in the case of a lower triangular matrix), which does not affect the determinant, gives $A \rightsquigarrow \mathrm{id}$, so that $\det A = 1$, so the claim holds in this case. □

Proposition 5.18. For $A \in \mathrm{Mat}_{n \times n}$, we have

$$\det A = \det(A^T),$$

i.e., the determinant does not change when passing from A to its transpose (Definition 4.88).

Proof. For small matrices (of size at most 3×3), this can be proved directly from the formulae in §5.2.1.

In general, one may argue like this: if A is not invertible, then A^T is not invertible either (by Lemma 4.90). In this case, both sides of the equation are zero. If A is invertible, it is a product of elementary matrices: $A = U_1 \dots U_n$. We then have $A^T = U_n^T \dots U_1^T$. By the product formula (Proposition 5.15), we may therefore assume A is an elementary matrix. In this case, one checks the claim by inspection:

- for $A = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$, we have $A^T = A$, so the claim clearly holds.

- Likewise, for $A = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & r & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$, we have $A = A^T$, so again the claim holds obviously.
- The matrix $\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$ is a lower triangular matrix, and its transpose an upper triangular matrix. Both have determinant 1 according to Proposition 5.17.

Remark 5.19. We introduced the determinant using *row* operations. The preceding result implies that one can replace the word "row" in all of the above by the word "column". Applying that, say, to Remark 5.9 we obtain that $\det A = 0$ if A has two identical columns.

Proposition 5.20. Let $A = (a_{ij}) \in \mathrm{Mat}_{n \times n}$. Then, for any i, one can compute the determinant using "*cofactor expansion*" along the i-th row. That is, the following identity holds, where c_{ij} are the cofactors of A (Definition 5.12):

$$\det A = a_{i1}c_{i1} + \cdots + a_{in}c_{in}.$$

Similarly, one can compute it using cofactor expansion along the j-th column, for any j:

$$\det A = a_{1j}c_{1j} + \cdots + a_{nj}c_{nj}.$$

For a proof of this, see, e.g. [Nic95, §3.6].

Example 5.21. For example, we expand the determinant along the second row:

$$\begin{aligned}
\det\begin{pmatrix} 2 & 3 & 7 \\ -4 & 0 & 6 \\ 1 & 5 & 0 \end{pmatrix} &= a_{21}c_{21} + a_{22}c_{22} + a_{23}c_{23} \\
&= (-4)(-1)^{1+2}\det\begin{pmatrix} 3 & 7 \\ 5 & 0 \end{pmatrix} + 0(-1)^{2+2}\det\begin{pmatrix} 2 & 7 \\ 1 & 0 \end{pmatrix} \\
&+ 6(-1)^{2+3}\det\begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix} \\
&= 4\cdot(-35) + 0 - 6\cdot 7 \\
&= -182.
\end{aligned}$$

The choice of the second row (as opposed to the others) is arbitrary, and the result is the same if we choose another row. However, the presence of the $a_{22} = 0$ simplifies the computation.

5.5 Exercises

Exercise 5.1. For which values of $a, b \in \mathbf{R}$ is the following matrix invertible? In this event, what is its inverse?

$$A = \begin{pmatrix} a & b & 3 \\ 2 & 1 & -1 \\ 1 & -1 & 4 \end{pmatrix}.$$

Exercise 5.2. Let A be a square matrix such that $A^2 = \mathrm{id}$ (the identity matrix). Prove that $\det(A) = \pm 1$.

Exercise 5.3. Compute the determinant of a *rotation matrix* (cf. Example 4.18),

$$A = \begin{pmatrix} \cos r & -\sin r \\ \sin r & \cos r \end{pmatrix}.$$

Exercise 5.4. Compute the determinant of

$$\begin{pmatrix} 2 & 0 & 1 & 4 \\ -1 & 3 & 0 & 2 \\ 1 & 0 & 2 & -3 \\ 0 & -2 & 5 & 1 \end{pmatrix}.$$

Exercise 5.5. (Solution at p. 261) Compute the determinant of $\begin{pmatrix} 3 & 0 & 0 \\ 1 & 4 & 0 \\ 2 & -3 & 5 \end{pmatrix}$ in three ways:

- by using Theorem 5.5,
- by using Sarrus's rule, (5.11),
- by using Proposition 5.17.

Exercise 5.6. (Solution at p. 261) Compute the determinants of

$$\begin{pmatrix} 1 & 5 & 8 \\ 40 & -9 & 1 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 5 & 8 \\ 40 & -9 & 1 \\ 1 & 5 & 8 \end{pmatrix}.$$

Exercise 5.7. Compute the determinants of

$$\begin{pmatrix} 3 & 26 & -9 & 3 \\ 0 & 3 & 1 & 28 \\ 0 & 0 & 2 & 71 \\ 0 & 0 & 0 & 3 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Exercise 5.8. Compute the inverses of

$$\begin{pmatrix} 10 & 9 \\ 11 & 10 \end{pmatrix} \text{ and } \begin{pmatrix} 5 & 2 & -1 \\ 0 & 0 & 1 \\ 6 & 2 & 3 \end{pmatrix}.$$

Chapter 6

Eigenvalues and eigenvectors

Eigenvalues and eigenvectors are an extremely useful concept of linear algebra. Coupled with certain numerical considerations, which are (only slightly!) beyond the scope of this course, eigenvalues have been used, for example, in the early PageRank algorithm employed by Google.

The overall idea of eigenvalues and eigenvectors is this: square matrices of the form

$$A = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

(i.e., the only non-zero entries are on the main diagonal; such matrices are called *diagonal matrices*) are particularly simple to comprehend and to use in computations. For example, products of the can be computed easily:

$$\begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & & 0 \\ & \ddots & \\ 0 & & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn}b_{nn} \end{pmatrix},$$

i.e., the (diagonal) entries can just be multiplied one by one, something that clearly goes wrong for products of general matrices. Eigenvalues and eigenvectors can, in certain cases, be used to reduce computations for general matrices to those for diagonal matrices.

6.1 Definitions

Definition 6.1. Let A be a *square* matrix. A real number λ is called an *eigenvalue* of A if there exists a *nonzero* vector $v \in \mathbf{R}^n$ that satisfies the equation

$$Av = \lambda v.$$

Such a vector v is called an *eigenvector* for the eigenvalue λ. In other words, multiplying the matrix A by the vector v results in a scaled version of v, where the scaling factor is the eigenvalue λ.

Likewise, for a linear map $f : V \to V$, λ is an eigenvalue if there is $v \in V, v \neq 0$ such that

$$f(v) = \lambda v. \tag{6.2}$$

We consider some of the linear maps $f : \mathbf{R}^2 \to \mathbf{R}^2$ of §4.2.1.

Example 6.3. If f is the reflection along, say, the x-axis, i.e., $f(x, y) = (x, -y)$, then (6.2) reads

$$(x, -y) = (\lambda x, \lambda y),$$

i.e., $x = \lambda x$ and $-y = \lambda y$. The first equation holds if $x = 0$ and λ arbitrary or if $\lambda = 1$ and x arbitrary. Similarly, the second forces $y = 0$ or $\lambda = -1$. Since, by definition, we have $(x, y) \neq (0, 0)$, we cannot have both $x = 0$ and $y = 0$. If $x \neq 0$, then $\lambda = 1$, which forces $y = 0$. If $x = 0$, then $y \neq 0$ and therefore $\lambda = -1$. Thus, there are two eigenvalues and eigenvectors are as follows:

- $\lambda = 1$, with eigenvectors $(x, 0)$ for arbitrary $x \in \mathbf{R}$,
- $\lambda = -1$, with eigenvectors $(0, y)$ for arbitrary $y \in \mathbf{R}$.

Before going on, we make the following observation, for any $n \times n$-matrix A. The equation

$$Av = \lambda v$$

can be rewritten as

$$0 = Av - \lambda v = Av - (\lambda \mathrm{id})v = (A - \lambda \mathrm{id})v.$$

Here we have used standard properties of matrix multiplication (Lemma 4.59). We seek a *non-zero* vector v satisfying this condition. Such a vector exists if and only if $A - \lambda\mathrm{id}$ is *not* invertible.

Example 6.4. We consider the shearing matrix $A = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$, where we assume $r \neq 0$ (otherwise $A = \mathrm{id}$). We check the invertibility of the matrix:

$$A - \lambda \mathrm{id} = \begin{pmatrix} 1-\lambda & r \\ 0 & 1-\lambda \end{pmatrix}.$$

It suffices to compute the determinant (Theorem 5.13):

$$\det \begin{pmatrix} 1-\lambda & r \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - r \cdot 0 = (\lambda - 1)^2.$$

This is zero precisely if $\lambda = 1$, i.e., $\lambda = 1$ is the only eigenvalue of A. Eigenvectors for this eigenvalue are those vectors such that

$$(A - 1\dot{\mathrm{id}}) \begin{pmatrix} x \\ y \end{pmatrix} = 0,$$

i.e., $\begin{pmatrix} ry \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & r \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$. Since we have assumed $r \neq 0$, this implies $y = 0$, and x is arbitrary. Thus, the eigenvectors for $\lambda = 1$ are of the form $\begin{pmatrix} x \\ 0 \end{pmatrix}$ for $x \in \mathbf{R}$.

6.2 The characteristic polynomial

Definition and Lemma 6.5. For $A \in \mathrm{Mat}_{n \times n}$, the function

$$\chi(t) = \det(A - t \cdot \mathrm{id}_n)$$

is a polynomial of degree n. It is called the *characteristic polynomial* of the matrix A. A real number λ is an eigenvalue of A if and only if

$$\chi(\lambda) = 0.$$

For example, for $A = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$, $\chi(t) = (t-1)^2$.

Example 6.6. We compute the eigenvalues of $A = \begin{pmatrix} 2 & -1 \\ 4 & -3 \end{pmatrix}$ using the characteristic polynomial:

$$\chi_A(t) = \det \begin{pmatrix} 2-t & -1 \\ 4 & -3-t \end{pmatrix} = (2-t)(-3-t) + 4 = t^2 + t - 2.$$

The equation $\chi_A(t) = 0$ solves as

$$t_{1/2} = -\frac{1}{2} \pm \sqrt{2 + \frac{1}{4}} = -\frac{1}{2} \pm \frac{3}{2},$$

i.e., the eigenvalues are $t_1 = -2$, $t_2 = 1$.

Example 6.7. We consider the rotation matrix $A = \begin{pmatrix} \cos r & -\sin r \\ \sin r & \cos r \end{pmatrix}$.

We have:

- if $r = \dots, -4\pi, -2\pi, 0, 2\pi, \dots$ (i.e., a rotation by a multiple of 360°, i.e., no rotation at all), then $A = \mathrm{id}_2$, and the only eigenvalue is $\lambda = 1$, and any vector $(x, y) \in \mathbf{R}^2$ is an eigenvector.
- if $r = \dots, -3\pi, -\pi, \pi, 3\pi, \dots$ (i.e., a rotation by 180° (plus an irrelevant multiple of 360°)), the only eigenvalue is $\lambda = -1$, and again any vector in $\mathbf{R}^2$ is an eigenvector,
- in all other cases, i.e., if the rotation is not by 0° or by 180°, there are *no* eigenvalues (and therefore no eigenvectors).

These statements are clear geometrically: for a rotation other than the special cases, for any vector $v \in \mathbf{R}^2$ we have that the rotated vector Av lies on a different line, so that it cannot be an eigenvector. To confirm this algebraically, we compute its characteristic polynomial:

$$\begin{aligned}
\chi(t) &= \det(A - t\mathrm{id}) \\
&= \det \begin{pmatrix} \cos r - t & -\sin r \\ \sin r & \cos r - t \end{pmatrix} \\
&= (\cos r - t)^2 + (\sin r)^2 \\
&= (\cos r)^2 - 2t \cos r + t^2 + (\sin r)^2 \\
&= 1 + t^2 - 2t \cos r.
\end{aligned}$$

We solve this for t using the usual formula:

$$t = 2 \cos r \pm \sqrt{-1 + (\cos r)^2}.$$

We have $0 \leqslant \cos r \leqslant 1$, and $\cos r = 1$ if and only if r is a multiple of π (cf. §B). Thus the term inside the square root is zero in this case, in all other cases it is strictly negative, so that the equation $\chi(t) = 0$ has *no* solution.

The non-existence of eigenvalues can be salvaged by working with complex numbers, instead of real numbers.

Theorem 6.8. (*Fundamental theorem of algebra*, cf. also §1.2 for further discussion) For every non-constant polynomial

$$p(t) = a_n x^n + \cdots + a_0,$$

where the coefficients $a_0, \ldots, a_n$ are complex numbers (for example, they can be real numbers), there exists a *complex* number $z \in \mathbf{C}$ such that

$$p(z) = 0.$$

This theorem is famous for the number of entirely different proofs. A completely elementary proof fitting on about two pages is given in [Oli11].

Example 6.9. Consider the rotation matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

describing a (counter-clockwise) rotation by 90°. Its characteristic polynomial

$$\chi_A(t) = 1 + t^2$$

does not have a real zero, i.e., for all real numbers r, $\chi_A(r) \neq 0$. However, the complex number i, the *imaginary unit*, which satisfies $i^2 = -1$ is a *complex* zero. In addition, $-i$, which also satisfies $(-i)^2 = -1$ is another complex zero.

It is therefore helpful to consider the concepts of linear algebra not only for real matrices, but for complex matrices. All of the concepts and theorems that we have encountered in this course continue to hold for complex matrices, complex vector spaces etc.

Corollary 6.10. Any complex square matrix $A \in \mathrm{Mat}_{n \times n}$ has at least one (complex) eigenvalue. In particular, any real square matrix has at least one complex eigenvalue (but it may not have a real eigenvalue).

6.3 Eigenspaces

In the above examples, the set of all eigenvectors for a given eigenvalue has a particularly nice shape. This is a general phenomenon:

Definition and Lemma 6.11. Let $A \in \mathrm{Mat}_{n\times n}$ be a square matrix and $\lambda \in \mathbf{R}$ a fixed real number. The set

$$E_\lambda := \{v \in \mathbf{R}^n \mid Av = \lambda v\}$$

is a subspace of $\mathbf{R}^n$. It is called the *eigenspace* of A with respect to λ.

Proof. The equation $Av = \lambda v$ is equivalent to $(A - \lambda \mathrm{id})v = 0$, i.e., we have $E_\lambda = \ker(A - \lambda \mathrm{id})$. This is a subspace of $\mathbf{R}^n$ by Proposition 4.23. □

Remark 6.12. If λ above is *not* an eigenvalue, then $E_\lambda = \{0\}$, i.e., the zero vector is the only one satisfying $Av = \lambda v$.

If λ is an eigenvalue, then E_λ consists of all the eigenvectors for the eigenvalue λ, together with the zero vector (which by definition is not an eigenvector).

Example 6.13. We compute the eigenspaces of the matrix $A = \begin{pmatrix} 0 & -1 \\ -2 & 0 \end{pmatrix}$. Its characteristic polynomial is

$$\chi_A(t) = \det \begin{pmatrix} -\lambda & -1 \\ -2 & -\lambda \end{pmatrix} = \lambda^2 - 2.$$

Its zeros, i.e., the eigenvalues of A are $\lambda_{1/2} = \pm\sqrt{2}$. The eigenspace for $\sqrt{2}$ is the solution space of the homogeneous system

$$\underbrace{\begin{pmatrix} -\sqrt{2} & -1 \\ -2 & -\sqrt{2} \end{pmatrix}}_{=A-\sqrt{2}\cdot\mathrm{id}=:B} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

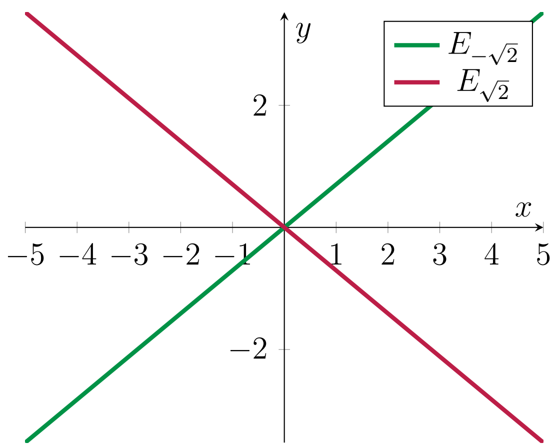
We solve this by reducing the matrix B to row-echelon form

$$\begin{pmatrix} -\sqrt{2} & -1 \\ -2 & -\sqrt{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ -2 & -\sqrt{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} \\ 0 & 0 \end{pmatrix}.$$

Thus, y is a free variable and $x = -\frac{\sqrt{2}}{2}y$. Thus $E_{\sqrt{2}}$ has dimension 1, a basis vector is $(-\frac{\sqrt{2}}{2}, 1)$. Similarly, one computes the eigenspace for $\lambda = -\sqrt{2}$:

$$B = A + \sqrt{2} \cdot \mathrm{id} = \begin{pmatrix} \sqrt{2} & -1 \\ -2 & \sqrt{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -\frac{1}{\sqrt{2}} \\ -2 & \sqrt{2} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -\frac{\sqrt{2}}{2} \\ 0 & 0 \end{pmatrix},$$

so the eigenspace $E_{-\sqrt{2}}$ is again one-dimensional, and a basis vector is $(\frac{\sqrt{2}}{2}, 1)$. Here is a plot showing the two eigenspaces: the map $v \mapsto Av$ will stretch the vectors in $E_{\sqrt{2}}$ by a factor of $\sqrt{2}$, while those on the eigenspace $E_{-\sqrt{2}}$ will be flipped and stretched by a factor of $\sqrt{2}$:



6.4 Diagonalizing matrices

As was mentioned above, diagonal matrices are particularly easy to compute with. This raises the question if (and how) it is possible to "bring" a given matrix A into such an easy form.

Definition 6.14. A square matrix A is called *diagonalizable* if there is an invertible matrix $P \in \mathrm{Mat}_{n \times n}$ such that

$$PAP^{-1} = D,$$

where $D = \begin{pmatrix} d_{11} & & 0 \\ & \ddots & \\ 0 & & d_{nn} \end{pmatrix}$ is a diagonal matrix.

An example showing the relevance of this notion is this: in the context of differential equations, one needs to compute the *exponential* of a square matrix A, which is defined as

$$\exp A = 1 + A + \frac{A^2}{2} + \frac{A^3}{6} + \frac{A^4}{24} + \dots .$$

Here $A^3 = A \cdot A \cdot A$ etc. Instead of computing all these powers of A one after another, one can use the above definition: if A is diagonalizable, i.e., $PAP^{-1} = D$, then $A = (P^{-1}P)A(P^{-1}P) = P^{-1}(PAP^{-1})P = P^{-1}DP$. Then,

$$A^2 = P^{-1}DP \cdot P^{-1}DP = P^{-1}D^2P, A^3 = P^{-1}D^3P$$

etc. Computing the powers of D, as opposed to those of A is easy: one just needs to raise the diagonal entries to the corresponding power.

Method 6.15. In order to diagonalize a square matrix $A \in \mathrm{Mat}_{n\times n}$, i.e., to determine whether P above exists, and to compute D, one proceeds as follows:

- Compute the eigenvalues of A, for example by finding the zeros of the characteristic polynomial. Denote them by $\lambda_1, \dots, \lambda_k$. Denote the eigenspaces by E_{λ_k}.

- The matrix A is diagonalizable precisely if

$$\sum_{i=1}^{k} \dim E_{\lambda_i} = n,$$

  i.e., if the dimensions of the eigenspaces sum up to the size of the matrix A.

- In this event, one may choose P to be the $n \times n$-matrix whose columns are the basis vectors of all the eigenspaces (for the various eigenvalues $\lambda_1, \dots, \lambda_k$). The matrix D is the diagonal matrix whose diagonal entries are

$$\underbrace{\lambda_1, \dots, \lambda_1}_{\dim E_1 \text{ times}}, \dots, \underbrace{\lambda_k, \dots, \lambda_k}_{\dim E_k \text{ times}}.$$

One can show that above, one always has $k \leqslant n$. One does this by proving that the sum of the subspaces $E_{\lambda_1}, \ldots, E_{\lambda_k}$ is a *direct sum*, so that

$$\begin{aligned} n = \dim \mathbf{R}^n &\geqslant \dim(E_{\lambda_1} + \cdots + E_{\lambda_k}) \\ &= \dim(E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}) \\ &= \dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_k}) \\ &\geqslant \underbrace{1 + \cdots + 1}_{k \text{ summands}} \\ &= k. \end{aligned}$$

This implies the following:

Corollary 6.16. If an $n \times n$-matrix has n distinct eigenvalues, then it is diagonalizable.

Definition 6.17. Let $A \in \mathrm{Mat}_{n \times n}$ be given. A basis $v_1, \ldots, v_n$ of $\mathbf{R}^n$ is called an *eigenbasis* for A if each v_i is an eigenvector (for a certain eigenvalue) of A.

Lemma 6.18. For $A \in \mathrm{Mat}_{n \times n}$ the following two statements are equivalent:
(1) A is diagonalizable.
(2) A admits an eigenbasis, i.e., there is an eigenbasis (of $\mathbf{R}^n$) for A.

One proves this by observing that if PAP^{-1} is a diagonal matrix, then P is a base-change matrix between the standard basis and an eigenbasis.

Example 6.19. The matrix $A = \begin{pmatrix} 0 & -1 \\ -2 & 0 \end{pmatrix}$ in Example 6.13 has two distinct eigenvalues, and is therefore diagonalizable (Corollary 6.16). An eigenbasis for A is

$$v_1 = (\sqrt{2}, 1), v_2 = (-\sqrt{2}, 1).$$

Example 6.20. We consider the *shearing matrix*

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Its characteristic polynomial is $\chi_A(t) = (1-t)^2$, whose only zero is $t = 1$. Thus, A has this eigenvalue only: $\lambda = 1$. We compute the eigenspace: consider the matrix $B := A - \lambda \mathrm{id} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Writing, as usual, $v = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbf{R}^2$, the space of solutions of the homogeneous system

$$Bv = \begin{pmatrix} y \\ 0 \end{pmatrix} = 0$$

is our eigenspace, namely

$$E_1 = \{v \in \mathbf{R}^2 \mid Bv = 0\} = \{(x, 0) \mid x \in \mathbf{R}\}.$$

This space is 1-dimensional, and has a basis consisting of the (single) vector $(1, 0)$. Thus, A is *not* diagonalizable.

Example 6.21. We continue the discussion of the rotation matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Its (complex) eigenvalues are $\lambda_1 = i$, $\lambda_2 = -i$. According to Corollary 6.16, A is diagonalizable. We compute the eigenspaces, where we regard A as a complex matrix:

$$E_i = \{v \in \mathbf{C}^2 \mid (A - i \cdot \mathrm{id})v = 0\}.$$

If $v = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, then

$$(A - i \cdot \mathrm{id})v = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} -iz_1 - z_2 \\ z_1 - iz_2 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This means $z_1 = iz_2$ from the second equation; the first is then also satisfied since $-iz_1 - z_2 = -i(iz_2) - z_2 = z_2 - z_2 = 0$. Thus

$$E_i = \{(iz, z) \mid z \in \mathbf{C}\},$$

i.e., as a complex vector space, E_i is 1-dimensional and a basis of it is the vector $(i, 1)$.

Similarly,

$$E_{-i} = \{v \in \mathbf{C}^2 \mid (A + i \cdot \mathrm{id})v = 0\}.$$

Computing this leads to the linear system

$$(A + i \cdot \mathrm{id})v = \begin{pmatrix} +i & -1 \\ 1 & +i \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} iz_1 - z_2 \\ z_1 + iz_2 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This gives $z_1 = -iz_2$, so that $E_{-i} = \{(-iz, z) \mid z \in \mathbf{C}\}$, and a basis of it is the (single) vector $(-i, 1)$. Thus, the matrix P above is

$$P = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}.$$

Example 6.22. For $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\lambda = 0$ is an eigenvalue (and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ an 0-eigenvector) and $\lambda = 1$ is another eigenvalue (and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ an 1-eigenvector).

6.5 Exercises

Exercise 6.1. Let $A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}$. Following Method 6.15, decide whether A is diagonalizable.

Help: you will find that the eigenvalues of A are among the numbers $0, 1, 2, 3$. You will be able to choose basis vectors of the eigenspaces all of whose coordinates are $-1, 0, 1$.

Exercise 6.2. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show that:

- $\chi_A(t) = t^2 - \mathrm{tr}(A)t + \det A = t^2 - (a + d)t + (ad - bc)$. Here $\mathrm{tr}(A)$ is the *trace* of A, cf. Exercise 4.24.
- The eigenvalues of A are

$$\lambda_{1/2} = \frac{a + d}{2} \pm \sqrt{\frac{(a - b)^2}{4} + 4bc}.$$

Exercise 6.3. For each of the following matrices, compute $\chi_A(t)$, the eigenvalues of A, the eigenspaces for these eigenvalues. Also decide whether A is diagonalizable and compute an eigenbasis if one exists.

(1) $A = \begin{pmatrix} 3 & 5 \\ 1 & -1 \end{pmatrix}$

(2) $A = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}$

(3) $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix}$

(4) $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

Exercise 6.4. Consider the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$. Compute its characteristic polynomial, its eigenvalues and its eigenspaces. Is A diagonalizable? If so, find a basis of $\mathbf{R}^3$ such that the associated matrix is a diagonal matrix, as in Definition 6.14.

Exercise 6.5. (Solution at p. 262) Let

$$f : \mathbf{R}^3 \to \mathbf{R}^3$$

be the linear map such that $f(1,0,1) = (2,0,2)$, $\ker f = L((1,1,1))$ and $f(2,0,-3) = (-2,0,3)$. Compute the matrix of f with respect to the standard basis.

Exercise 6.6. For which $a \in \mathbf{R}$ is the matrix

$$A_a = \begin{pmatrix} a & 0 & 0 \\ a-2 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

diagonalizable?

Exercise 6.7. (Solution at p. 263) For a parameter $a \in \mathbf{R}$, let

$$A_a = \begin{pmatrix} 4 & 0 & 4 \\ a & 2 & a \\ -2 & 0 & -2 \end{pmatrix}.$$

(1) Compute the characteristic polynomial and the eigenvalues of A_a, for all $a \in \mathbf{R}$.

(2) Compute the values of a for which A_a is diagonalizable. For these a, find an invertible matrix P such that $P^{-1}A_aP$ is a diagonal matrix.

Exercise 6.8. Consider the matrices

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(1) Compute the eigenvalues and eigenvectors of A and show A is diagonalizable.

(2) Show that the characteristic polynomials of A and B are the same. Compute the eigenvalues and eigenspaces of B. Explain why A and B do *not* represent the same linear map with respect to different bases!

Exercise 6.9. For a parameter $t \in \mathbf{R}$, consider the matrix

$$A_t = \begin{pmatrix} -1 & 2 & t \\ 2 & 0 & -2 \\ t & -2 & -1 \end{pmatrix}.$$

(1) For which values of t does A_t have 0 as an eigenvalue?

(2) Compute the eigenvalues and eigenspaces of A_t for those values of t obtained in the previous part.

Exercise 6.10. Consider the space $\mathrm{Mat}_{2\times 2}$ of 2×2-vector spaces. Consider $A = \begin{pmatrix} -4 & 8 \\ 1 & -2 \end{pmatrix}$ and the linear map

$$F : \mathrm{Mat}_{2\times 2} \to \mathrm{Mat}_{2\times 2}, X \mapsto AX.$$

(1) Compute the 4×4-matrix of F with respect to the standard basis of $\mathrm{Mat}_{2\times 2}$, i.e., the matrices

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

(2) Compute a basis of $\ker F$ and $\operatorname{im} F$.

(3) Compute the eigenvalues and eigenspaces of F.

Remark 6.23. The linearity of F is a consequence of Lemma 4.59. It is also very similar to Proposition 4.19.

Exercise 6.11. (Solution at p. 264) Consider the two matrices

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 2 \\ 0 & 0 & 3 \end{pmatrix}.$$

Do they represent the same linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ (with respect to different bases)?

Exercise 6.12. (Solution at p. 265) Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$. Determine the eigenvalues of A and the corresponding eigenspaces. Is A diagonalizable? Is A^2 *similar* to A? I.e., does A^2 represent the same linear map $\mathbf{R}^3 \to \mathbf{R}^3$ as A?

Exercise 6.13. (Solution at p. 265) Consider the vectors $v_1 = (1, 0, 1)$, $v_2 = (1, 1, 1)$ and $v_3 = (1, 1, 2)$.

(1) Explain why there is a unique linear map $f : \mathbf{R}^3 \to \mathbf{R}^3$ such that $f(v_1) = (0, 0, 0)$, $f(v_2) = (1, 0, 3)$ and v_3 is an eigenvector of eigenvalue 4.

(2) Compute the matrix A of f with respect to the basis v_1, v_2, v_3 (both on the domain and on the codomain).

(3) Compute the matrix B of f with respect to the standard basis (both for the domain and the codomain).

(4) For $t \in \mathbf{R}$, consider the vector $v_t = (2, t, 5)$. For which values of t is $v_t \in \operatorname{im} f$?

Exercise 6.14. (Solution at p. 267) Consider the matrix

$$A = \begin{pmatrix} 0 & 2 & t \\ -3 & -5 & 6 \\ -2 & -2 & 5 \end{pmatrix}$$

(1) Determine the value of t for which A is *not* invertible.

(2) We now put $t = 2$ for the remainder of this exercise. Determine the value of a for which the vector $v = (2, 0, a)$ is an eigenvector of A. What is the corresponding eigenvalue?

(3) Determine all the eigenvalues of A and decide whether A is diagonalizable.

(4) Decide whether A is similar to the matrix A^2 (justify your response).

Exercise 6.15. (Solution at p. 268) Consider the matrix

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 1 & 1 & 2 \\ -1 & 1 & t \end{pmatrix}$$

(1) Compute the value of t for which the kernel of A is different from $\{0\}$.

(2) For the remainder of the exercise we put t to be equal to the value computed in part (a). Compute the characteristic polynomial and the eigenvalues of A.

(3) Find an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.

(4) Explain why any 3×3-matrix B such that $\chi_A(t) = \chi_B(t)$ is diagonalizable.

Exercise 6.16. (Solution at p. 268) Consider the matrix $A = \begin{pmatrix} 1 & 0 & t \\ 1 & 2 & 1 \\ 2 & 0 & -1 \end{pmatrix}$

(1) For what value of $t \in \mathbf{R}$ is the matrix A *non*-invertible?

(2) For each $t \in \mathbf{R}$, determine the eigenvalues of A. Specify for which values $t \in \mathbf{R}$ all the eigenvalues of A are real numbers.

(3) Determine for which $t \in \mathbf{R}$ there are eigenvalues with multiplicity > 1.

(4) For the value of t found in the first part of the exercise: compute a basis of all eigenspaces and decide whether A is diagonalizable.

Chapter 7

Euclidean spaces

The definition of a (real) vector space encodes the existence (and good properties) of the addition of vectors and the scalar multiplication of vectors. The vector space $\mathbf{R}^n$ has, however, another important piece of structure, namely the distance between two points, and the property of vectors being orthogonal to each other.

7.1 The scalar product on $\mathbf{R}^n$

Definition 7.1. The *scalar product* of $v, w \in \mathbf{R}^n$ is defined as

$$\langle v, w \rangle := v^T \mathrm{id} w = v^T w = v_1 w_1 + \cdots + v_n w_n.$$

(This is not to be confused with the scalar multiple of a vector, which is again a vector!)

Example 7.2. The scalar product can be positive, zero, or negative:

- $\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix} \right\rangle = 1 \cdot (-2) + 2 \cdot 2 = 2$
- $\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\rangle = 1 \cdot (-2) + 2 \cdot 1 = 0$
- $\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \end{pmatrix} \right\rangle = 1 \cdot (-2) + 2 \cdot 0 = -2$

However, for any $v \in \mathbf{R}^n$, we have

$$\langle v, v \rangle = \sum_{i=1}^{n} v_i^2 \geqslant 0 \tag{7.3}$$

i.e., a scalar product of a vector with *itself* is always non-negative. This implies that

$$\|v\| := \sqrt{\langle v, v\rangle} = \sqrt{v_1^2 + \cdots + v_n^2}$$

is a well-defined (real) number. It is called the *norm* of the vector v.

Lemma 7.4. The norm $\|v\|$ is the length of the line segment from the origin to v.

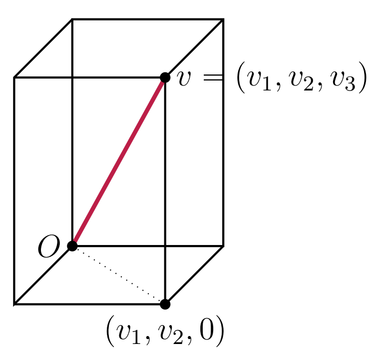
For $v, w \in \mathbf{R}^2$, there holds

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\| \cos r,$$

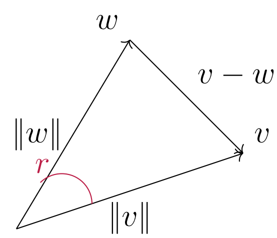
where r is the angle between the vector v and w.

Proof. The formula for the norm follows from repeatedly applying the *Pythagorean theorem*. Illustrating this for $n = 3$, we see that the line segment (shown dotted below) from the origin $O = (0, 0, 0)$ to the point $(v_1, v_2, 0)$ has length $\sqrt{v_1^2 + v_2^2}$. Therefore the length of the segment from O to v is

$$\sqrt{\left(\sqrt{v_1^2 + v_2^2}\right)^2 + v_3^2} = \sqrt{v_1^2 + v_2^2 + v_3^2}.$$



The formula for the norm of $v - w$ follows is a reformulation of the *law of cosines*.



□

Given a square matrix $A \in \mathrm{Mat}_{n\times n}$, we have considered so far the linear map

$$\mathbf{R}^n \to \mathbf{R}^n, v \mapsto A \cdot v.$$

In addition to that, there is another fundamental map that one can associate to a matrix:

$$\langle -, - \rangle_A : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}, (v, w) \mapsto \langle v, w \rangle_A := v^T \cdot A \cdot w.$$

Here we regard v and w as column vectors, i.e., as $n \times 1$-matrices. Therefore, for $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, $v^T = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix}$ is a row vector (with n entries). Therefore $v^T A$ is an $1 \times n$-matrix, so that $v^T Aw$ is an 1×1-matrix, i.e., just a real number. We call this number the *scalar product* of v and w with respect to the given matrix A.

Lemma 7.5. The scalar product has the following fundamental properties:

- If we fix $w \in \mathbf{R}^n$, then the maps

$$\begin{aligned} \langle ?, w \rangle : &\mathbf{R}^n \to \mathbf{R}, v \mapsto \langle v, w \rangle \\ \langle w, ? \rangle : &\mathbf{R}^n \to \mathbf{R}, v \mapsto \langle w, v \rangle \end{aligned}$$

  are linear (cf. Definition 4.1; e.g., for the first this means concretely that

$$\langle rv + r'v', w \rangle = r\langle v, w \rangle + r'\langle v', w \rangle,$$

  for $r, r' \in \mathbf{R}$, $v, v' \in \mathbf{R}^n$. We refer to this by saying that $\langle -, - \rangle : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ is a *bilinear form* (or as the *bilinearity* of the scalar product).

- We have

$$\langle v, w \rangle = \langle w, v \rangle.$$

  This property is called *symmetry*.

Proof. By Proposition 4.19, the map $w \mapsto v^T w = \langle v, w\rangle$ is linear. The proof of the linearity in the first argument is similar, or it follows from symmetry.

The identity $\langle v, w\rangle = \langle w, v\rangle$ is directly clear from the definition. One may also prove it using (4.91):

$$(v^T w)^T = w^T (v^T)^T = w^T v.$$

Noting that any 1×1-matrix (such as $v^T w$) is equal to its transpose, the left hand side equals $\langle v, w\rangle$, while the right equals $\langle w, v\rangle$. □

Using the bilinearity of $\langle -, -\rangle$, we can compute the following expression

$$\begin{aligned}\|v - w\|^2 &= \langle v - w, v - w\rangle \\ &= \langle v, v - w\rangle - \langle w, v - w\rangle \\ &= \langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle \\ &= \|v\|^2 + \|w\|^2 - 2\langle v, w\rangle.\end{aligned}$$

Comparing this with the cosine law above we see

$$\langle v, w\rangle = \|v\|\|w\| \cos r.$$

The factor $\cos r$ is equal to 0 precisely if $r = -\frac{\pi}{2}, \frac{\pi}{2}$ (i.e., 90° or -90°). In other words,

$$\langle v, w\rangle = 0$$

if the angle between the vectors v and w is $\pm 90^\circ$. This motivates the following definition.

Definition 7.6. Two vectors $v, w \in \mathbf{R}^n$ are said to be *orthogonal* if

$$\langle v, w\rangle = \sum_{i=1}^{n} v_i w_i = 0.$$

7.2 Positive definite matrices

Definition and Lemma 7.7. If A is a *symmetric* $n \times n$-matrix (i.e., $A = A^T$), then the map

$$\langle -, -\rangle_A : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}, \langle v, w\rangle_A := v^T A w$$

is bilinear and symmetric, i.e., Lemma 7.5 holds verbatim for $\langle -, -\rangle_A$ instead of the standard scalar product (which corresponds to the case $A = \mathrm{id}_n$).

Example 7.8. Suppose $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$. Then $Aw = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ -w_4 \end{pmatrix}$, so that

$$\begin{aligned} \langle v, w \rangle_A = v^T A w &= \begin{pmatrix} v_1 & v_2 & v_3 & v_4 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ -w_4 \end{pmatrix} \\ &= v_1 w_1 + v_2 w_2 + v_3 w_3 - v_4 w_4. \end{aligned}$$

This example is not an anomaly, but the basis of so-called *Minkowski space* which is fundamental in special relativity, which is $\mathbf{R}^{3+1}$ with 3 space coordinates and 1 time coordinate.

The standard basis vectors $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \dots, e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ are orthogonal to each other, but

$$\langle e_4, e_4 \rangle_A = -1$$

where as $\langle e_k, e_k \rangle_A = +1$ for the other three basis vectors. In that sense, the scalar product (with respect to A) is able to distinguish between the last and the other three directions.

Definition 7.9. A symmetric matrix A is called *positive definite* if

$$\langle v, v \rangle_A > 0$$

for all $v \in \mathbf{R}^n$, $v \neq 0$. In this case we can define the *norm* (of v with respect to the matrix A) as

$$\|v\|_A := \sqrt{\langle v, v \rangle_A}.$$

It is *negative definite* if instead $\langle v, v \rangle_A < 0$ for all $v \neq 0$. The matrix A is called *indefinite* if there exist $v, w \in \mathbf{R}^n$ with $\langle v, v \rangle_A > 0$ and $\langle w, w \rangle_A < 0$.

Example 7.10. As we have seen in (7.3), id_n is positive definite. The matrix in Example 7.8 is indefinite.

It is suggestive to blame the -1 in the last entry for the indefiniteness of the matrix in Example 7.8. The following result gives a way to ensure positive definiteness for general matrices. To state it, we introduce a bit of terminology:

Definition 7.11. For a square matrix A, the *principal submatrix* (of size r) is the matrix

$$A^{(r)} = (a_{ij})_{1 \leqslant i,j \leqslant r}.$$

I.e., it is the matrix consisting of the first r rows and columns of A.

Proposition 7.12. Let $A \in \mathrm{Mat}_{n \times n}$ be a symmetric square matrix. The following are equivalent:

(1) the bilinear form $\langle -, - \rangle_A$ is positive definite, i.e., $\langle v, v \rangle_A \geqslant 0$ for all $v \in \mathbf{R}^n$,

(2) A is positive definite,

(3) For all $1 \leqslant r \leqslant n$, $\det(A^{(r)}) > 0$.

In particular, any positive definite matrix A has $\det A > 0$. Therefore such a matrix is invertible (Theorem 5.13).

A proof of this criterion requires methods from §7.3.

Example 7.13. Consider the matrix $A = \begin{pmatrix} 1 & 2 & t \\ 2 & 5 & 8 \\ t & 7 & 14 \end{pmatrix}$, where $t \in \mathbf{R}$ is some parameter. We inspect its positive definiteness: since $A^{(1)} = 1$ is positive, $\det A^{(2)} = \det \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} = 1 > 0$ and $\det A = \det A^{(3)} = -5t^2 + 32t - 50$. For $t = 3$, this equals $+1$, so the matrix A is positive definite in this case. For $t = 4$, this equals -2, so the matrix A is indefinite in this case.

Example 7.14. The defininiteness of matrices has applications in analysis: for a (twice differentiable) function $f : \mathbf{R}^2 \to \mathbf{R}$, such as

$$f(x, y) = x^2 + y^2,$$

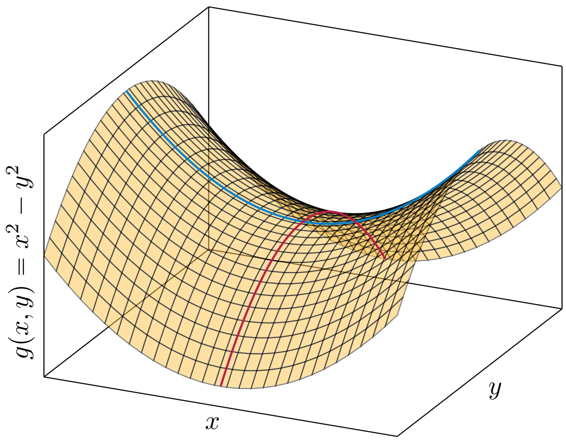


Figure 7.1: The function $g(x,y) = x^2 - y^2$ has a *saddle point* at $(0,0)$; informally this means that there are directions in which g increases (here the x-direction, shown blue), and directions in which g decreases (the y-direction, red parabola).

one considers the so-called *Hesse matrix*, which is given by

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial x y} \end{pmatrix}.$$

For the above function it is

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

which is positive definite. By contrast, for $g(x,y) = x^2 - y^2$, it is $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$, which is indefinite. One proves in analysis that the positive defininetess of the Hesse matrix implies that there is a local minimum at a given point (x,y), provided that $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$ at this point. Thus, f has a local minimum at the point $(0,0)$, but g does not.

7.3 Euclidean spaces

Definition 7.15. A *Euclidean vector space* is a vector space V together with a map

$$\langle -, - \rangle : V \times V \to \mathbf{R}$$

that is

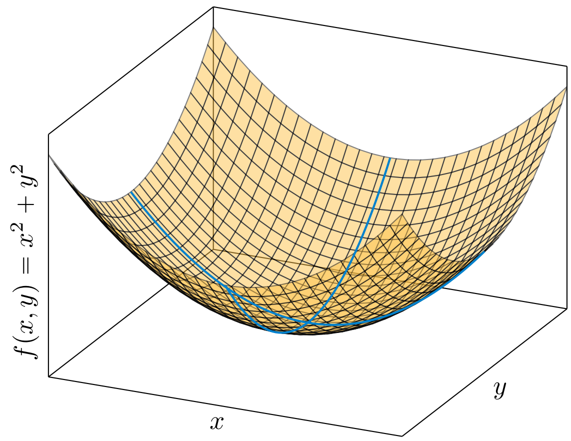


Figure 7.2: The function $f(x, y) = x^2 + y^2$ has a local minimum at $(0, 0)$. Informally this means that moving in any direction from the point $(0, 0)$, the value of $f(x, y)$ increases.

- bilinear (i.e., $\langle v, - \rangle$ and $\langle -, v \rangle : V \to \mathbf{R}$ are linear for each $v \in V$),
- symmetric (i.e., $\langle v, w \rangle = \langle w, v \rangle$), and
- positive definite ($\langle v, v \rangle > 0$ for each $v \neq 0$).

One also refers to the map $\langle -, - \rangle$ as the scalar product on V. We say v, w are *orthogonal* if $\langle v, w \rangle = 0$. We will indicate this by writing

$$v \perp w.$$

We call

$$\|v\| := \sqrt{\langle v, v \rangle} (\in \mathbf{R}^{\geqslant 0})$$

the norm of the vector v. For $v, w \in V$, the *distance* between v and w is defined as

$$d(v, w) := \|v - w\|.$$

Example 7.16. (1) $\mathbf{R}^n$ with the above scalar product is an Euclidean vector space. More generally, for a symmetric, positive definite matrix A, $\mathbf{R}^n$ together $\langle -, - \rangle_A$ is an Euclidean space.

In other words, the above turns the fundamental properties of $\mathbf{R}^n$, together with the standard scalar product (or, more generally $\mathbf{R}^n$ with the scalar product $\langle -, - \rangle_A$ given by a positive definite symmetric matrix A) into an abstrac definition, similarly to the way that a vector space is an abstraction of the key properties of $\mathbf{R}^n$.

(2) If V, together with some given scalar product $\langle -, - \rangle$ is a Euclidean space, then so is any subspace of V. In particular, any subspace of $\mathbf{R}^n$ with the standard scalar product is again an Euclidean space. For example, any plane inside $\mathbf{R}^3$ is an Euclidean space.

(3) One can use elementary properties of the integral to show that the vector space $C = C([-1, 1])$ of continuous functions $f : [-1, 1] \to \mathbf{R}$ with

$$\langle f, g \rangle := \int_{-1}^{1} f(x)g(x)dx$$

is an (infinite-dimensional) Euclidean space, which is of fundamental importance in analysis.

(4) As in Example 7.8, consider again $V = \mathbf{R}^n$, but

$$\langle v, w \rangle := v_1 w_1 + \cdots + v_{n-1} w_{n-1} - v_n w_n.$$

This is bilinear and symmetric, but *not* positive definite, and therefore not a scalar product.

Proposition 7.17. Let $(V, \langle -, - \rangle)$ be an Euclidean space. For each $v, w \in V$, there holds:

(1) $\|v\| \geqslant 0$,

(2) $\|v\| = 0$ if and only if $v = 0$,

(3) $\|rv\| = |r|\|v\|$ for $r \in \mathbf{R}$,

Proof. The first and third statement is immediate. The second holds since $\langle -, - \rangle$ is (by definition) positive definite. □

The scalar product yields a crucial additional feature that general vector spaces do not possess. This is based on the following idea. Throughout, let $(V, \langle -, - \rangle)$ be an Euclidean vector space.

Lemma 7.18. Let $e \in V$ be a vector of norm 1, i.e., $\|e\| = 1$. Let $v \in V$ be any vector. Then the vector

$$\tilde{v} := v - \langle v, e \rangle \cdot e$$

is orthogonal to e and we have the equation

$$v = \tilde{v} + \langle v, e\rangle \cdot e \tag{7.19}$$

expressing v as a sum of a scalar multiple of e and a vector that is orthogonal to e.

Proof. The orthogonality of $\tilde{v}$ and $\langle v, e\rangle \cdot e$ is a computation using the bilinearity of $\langle -, -\rangle$:

$$\begin{aligned}\langle \tilde{v}, \langle v, e\rangle \cdot e\rangle &= \langle v - \langle v, e\rangle \cdot e, \langle v, e\rangle \cdot e\rangle \\ &= \langle v, e\rangle(\langle v - \langle v, e\rangle \cdot e, e\rangle) \\ &= \langle v, e\rangle(\langle v, e\rangle - \langle\langle v, e\rangle \cdot e, e\rangle) \\ &= \langle v, e\rangle(\langle v, e\rangle - \langle v, e\rangle \underbrace{\langle e, e\rangle}_{=1}) \\ &= 0.\end{aligned}$$

The equation (7.19) is obvious from the definition of $\tilde{v}$. □

We now extend the observation of Lemma 7.18 to more than a single vector. To do so, we introduce some terminology.

Definition and Lemma 7.20. The *orthogonal complement* of a subset $M \subset V$ is defined as

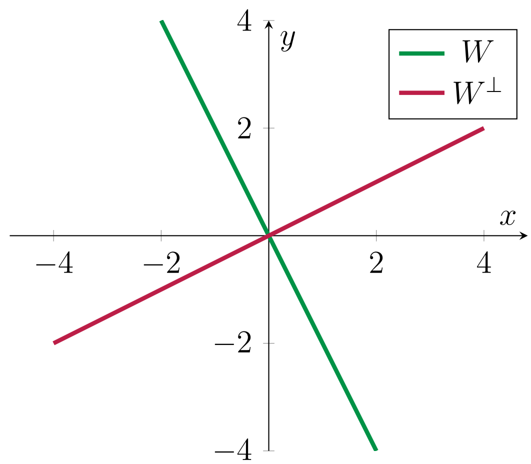
$$M^\perp := \{v \in V \mid \langle v, m\rangle = 0 \text{ for all } m \in M\}.$$

This is a sub*space* of V.

For a subspace W, one has

$$W \cap W^\perp = \{0\}, \qquad (7.21).$$

The last assertion can be rephrased by saying that the zero vector is the only element in W that is orthogonal to all vectors in W. Colloquially, this means that if W gets larger, then $W^\perp$ gets smaller. This idea is made more precise (in terms of dimensions) in Corollary 7.30 below. The last assertion is proved using the positive-definiteness of $\langle -, -\rangle$ (specifically, Proposition 7.17(2)).



Example 7.22. Consider the subspace $W = L(\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}) \subset \mathbf{R}^3$ (with its standard scalar product). We compute $W^\perp$. A vector $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbf{R}^3$ will be orthogonal to W if and only if it is orthogonal to $v_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. This follows from the linearity of $\langle x, - \rangle$. We make the conditions $x \perp v_1$ and $x \perp v_2$ explicit:

$$\begin{aligned} x \perp v_1 &\Rightarrow x_1 + 2x_2 + 4x_3 &= 0 \\ x \perp v_2 &\Rightarrow x_1 + x_2 &= 0. \end{aligned}$$

We solve this homogeneous system

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 4 \\ 0 & -1 & -4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \end{pmatrix}$$

which shows that x_3 is a free variable, and that the solution space of the system, i.e., $W^\perp$ is the subspace

$$W^\perp = L(\begin{pmatrix} 4 \\ -4 \\ 1 \end{pmatrix}).$$

Definition 7.23. A family $v_1, \dots, v_n$ of vectors is called an *orthonormal system* if

- $\|v_i\| = 1$ (i.e., $\langle v_i, v_i \rangle = 1$) for all i,
- $v_i \perp v_j$ (i.e., $\langle v_i, v_j \rangle = 0$) for all $i \neq j$.

If the vectors additionally form a basis of V, then we speak of an *orthonormal basis*.

For example, the standard basis in $\mathbf{R}^n$ is an orthonormal basis (with respect to the standard scalar product).

Theorem 7.24. Let $u_1, \dots, u_n$ be an orthonormal system (in an Euclidean space). Let $U = L(u_1, \dots, u_n) \subset V$ be the subspace spanned by these vectors. Then there is a unique linear map, called the *orthogonal projection*

$$p : V \to U$$

such that
(1) $p(u) = u$ for all $u \in U$,

(2) $p(v) - v \in U^\perp$ for all $v \in V$.
In particular, every vector $v \in V$ can be written as

$$v = \underbrace{p(v)}_{\in U} + \underbrace{v - p(v)}_{\in U^\perp},$$

i.e., a sum of a vector in U and another one in its orthogonal complement $U^\perp$. This is the unique representation of v in such a form.

The map p is given by

$$p(v) = \sum_{k=1}^{n} \langle v, u_k \rangle u_k. \tag{7.25}$$

Example 7.26. In $V = \mathbf{R}^3$, equipped with its standard scalar product, we consider $u_1 = (1, 0, 0)$ and $u_2 = (0, 1, 0)$. These form an orthonormal system. Then $U = L(u_1, u_2) = \{(x, y, 0) | x, y \in \mathbf{R}\}$ is the x-y-plane; its orthogonal complement is $U^\perp = L((0, 0, 1)) = \{(0, 0, z) | z \in \mathbf{R}\}$, the z-axis. The orthogonal projection as defined in (7.25) sends a vector $v = (x, y, z)$ to

$$p(v) = \langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 = x u_1 + y u_2 = (x, y, 0).$$

Proof. The map p defined in (7.25) is linear, since $\langle -, u_k \rangle$ is linear. It satisfies the two conditions. One checks this using that the u_k form an orthonormal system, very similarly to the proof of Lemma 7.18.

If $q : V \to U$ is another map with these two properties, we have

$$\langle p(v) - q(v), u \rangle = \left\langle \underbrace{p(v) - v - (q(v) - v)}_{\in U^\perp}, u \right\rangle = 0$$

for all $v \in V$, $u \in U$. Since $q(v)$, $p(v) \in U$, we have $p(v) - q(v) \in U$. Thus, the vector $p(v) - q(v)$ is zero, by (7.21). This shows the unicity of p.

The final claim holds since $v = \underbrace{p(v)}_{\in U} + \underbrace{v - p(v)}_{\in U^\perp}$ is such a representation. If $v = u_1 + u'_1$ with $u_1 \in U$ and $u'_1 \in U^\perp$ is another such representation, then $u - u_1 = u' - u'_1$ lies both in U (left hand side), but also in $U^\perp$ (right hand side). However, again applying Proposition 7.17(2) to U, we have $U \cap U^\perp = \{0\}$, so $u = u_1$ and $u' = u'_1$. □

Corollary 7.27. Suppose $u_1, \dots, u_n$ form an orthonormal system (of a Euclidean vector space $(V, \langle -, - \rangle)$) such that V is spanned by these vectors. Then

- the following formula holds for any $v \in V$:

$$v = \sum_{k=1}^{n} \langle v, u_i \rangle u_i. \tag{7.28}$$

- The vectors are necessarily linearly independent, i.e., they form an orthonormal *basis*.

Proof. We apply Theorem 7.24 to these vectors. By the assumption $U = V$, so that by (1), $p(v) = \text{id}$. The first claim then holds by (7.25).

If $0 = \sum_{k=1}^n a_k u_k$ is a linear combination, we apply $\langle -, u_l \rangle$, for any $1 \leqslant l \leqslant n$, to (7.28):

$$\begin{aligned} 0 &= \langle 0, u_l \rangle \\ &= \left\langle \sum_{k=1}^{n} a_k u_k, u_l \right\rangle \\ &= \sum_{k=1}^{n} a_k \langle u_k, u_l \rangle. \end{aligned}$$

In this sum, all terms except the one with $k = l$ are zero, since $u_k \perp u_l$ for $k \neq l$. We also have $\langle u_l, u_l \rangle = 1$, which shows that $a_l = 0$, and therefore the linear independence of the given vectors. □

Example 7.29. The standard basis $e_1, \dots, e_n$ of $\mathbf{R}^n$ is an orthonormal basis. For $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, we have $\langle e_i, v \rangle = v_i$ and the representation in (7.28) is the usual expansion of v:

$$v = v_1 e_1 + \dots + v_n e_n.$$

In general, the identity (7.28) is a convenient way to compute the coordinates of a given vector in terms of an (orthonomal) basis.

Using these results, one can quickly prove:

Corollary 7.30. If $U \subset V$ is a subspace of a finite-dimensional Euclidean space then

$$\dim U^\perp = \dim V - \dim U.$$

The presence of a positive definite (symmetric) matrix yields the following algorithmic device that constructs a particularly convenient set of vectors.

Proposition 7.31. (*Gram–Schmidt orthogonalization*) Let $v_1, \dots, v_r$ be any set of linearly independent vectors (in an Euclidean space). Then the vectors $w_1, \dots, w_r$ defined inductively as follows are an orthonormal system: They are constructed as follows

$$\begin{aligned}
w_1 &:= \frac{1}{\|v_1\|} v_1 && \text{(normalization)} \\
w_2' &:= v_2 - \langle v_2, w_1 \rangle w_1 && \text{(orthogonalization w.r.t. } L(w_1)) \\
w_2 &:= \frac{1}{\|w_2'\|} w_2' && \text{(normalization)} \\
&\vdots \\
w_r' &:= v_r - \sum_{k=1}^{r-1} \langle v_r, w_k \rangle \cdot w_k && \text{(orthogonalization w.r.t. } L(w_1, \dots, w_{r-1})) \\
w_r &:= \frac{1}{\|w_r'\|} w_r' && \text{(normalization)}
\end{aligned}$$

We have

$$L(v_1, \dots, v_r) = L(w_1, \dots, w_r).$$

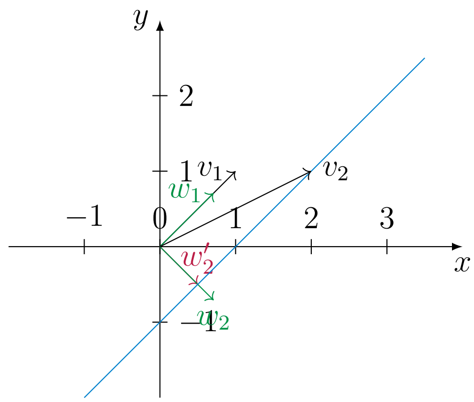
In particular, if the v_i form a basis, then so do the w_i, i.e., they then form an orthonormal basis. Yet more in particular, this shows that any finite-dimensional Euclidean space admits an orthonormal basis.

Proof. In each step, the vector w'_r is constructed in such a way that w'_r is orthogonal to the preceding vectors $w_1, \dots, w_{r-1}$, cf. (7.25). The division by the norms of the vectors w'_r ensures that $\|w_r\| = 1$. Note that this is possible since $\|w'_r\| > 0$ since $w'_r \neq 0$ and $\langle -, - \rangle$ is positive definite. □

Example 7.32. We consider $A = \mathrm{id}_2$, i.e., the standard scalar product on $\mathbf{R}^2$ and $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. (One checks this is a basis of $\mathbf{R}^2$!) Then

$$\begin{aligned} w_1 &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ w'_2 &= v_2 - \langle v_2, w_1 \rangle w_1 \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \frac{3}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ w_2 &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned}$$

Here is an illustration of the method in this example. The blue line depicts the vectors of the form $v_2 + aw_1$ for $a \in \mathbf{R}$. The vector w'_2 is the vector on that line that is orthogonal to w_1:



Corollary 7.33. Let $U \subset V$ be a subspace of a finite-dimensional Euclidean space V. Then there are two unique linear maps, called the *orthogonal projection* onto U, resp. onto $U^\perp$,

$$p_U : V \to U$$
$$p_{U^\perp} : V \to U^\perp$$

such that every vector $v \in V$ can be written as

$$v = p_U(v) + p_{U^\perp}(v). \tag{7.34}$$

Proof. By Proposition 7.31, U has an orthonormal basis, so we can apply Theorem 7.24, which gives us the orthogonal projection $p_U : V \to U$. If we define $p_{U^\perp}(v) := v - p_U(v)$, (7.34) holds by design, moreover, $p_{U^\perp}(v) \in U^\perp$ again by Theorem 7.24. The unicity of a decomposition as in (7.34) is again part of Theorem 7.24. □

7.4 Orthogonal and symmetric matrices

Definition 7.35. A real square matrix $A \in \mathrm{Mat}_{n\times n}$ is called *orthogonal* if

$$AA^T = \mathrm{id}.$$

This is equivalent to saying that A is invertible and $A^{-1} = A^T$. The following lemma explains the name “orthogonal”.

Lemma 7.36. For a square matrix $A \in \mathrm{Mat}_{n\times n}$, the following are equivalent:
(1) A is orthogonal,

(2) the n rows are an orthonormal basis of $\mathbf{R}^n$,

(3) the n columns are an orthonormal basis of $\mathbf{R}^n$.

Proof. If e_i is the i-th standard basis vector, we know that Ae_i is the i-th column A. We compute

$$\langle Ae_i, Ae_j \rangle = (Ae_i)^T (Ae_j) = e_i^T A^T A e_j.$$

The vector $A^T A e_j$ is the j-th column of $A^T A$, and the number $e_i^T A^T A e_j$ is the i-th entry of that vector. Thus, saying that the above expression equals 1 for $i = j$ and 0 otherwise is equivalent to requiring $A^T A = \mathrm{id}$. □

Theorem 7.37. The following conditions are equivalent for an $n \times n$-matrix A:

(1) A is symmetric,

(2) A is *orthogonally diagonalizable*, i.e., there is an *orthogonal matrix* P such that $P^{-1}AP$ is a diagonal matrix,

(3) A has an orthonormal eigenbasis.

If these equivalent conditions hold, then the columns of P form an orthonormal eigenbasis and vice versa. (Note that $P^{-1} = P^T$ can be computed without computing, properly speaking, the inverse of P.)

The implication (1) $\Rightarrow$ (2) in particular says:

$$A \text{ symmetric} \Rightarrow A \text{ diagonalizable.}$$

For a proof of this theorem, see, e.g. [Nic95, Theorem 8.2.2]. The vectors of an orthonormal eigenbasis are also called the *principal axes* of A. The theorem is sometimes called the *principal axes theorem*. We only point out that the difficult direction is to show that (1) $\Rightarrow$ (2). One does this by proving that a *symmetric* real matrix has only real eigenvalues (as opposed to complex). For 2×2-matrices, one can see this by direct computation (see also Exercise 6.2): the characteristic polynomial of a symmetric 2×2-matrix $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ is

$$\chi_A(t) = \det(A - t\,\mathrm{id}) = (a-t)(d-t) - b^2 = t^2 + (-a-d)t + ad - b^2.$$

The zeroes of this polynomial are given by

$$\begin{aligned}\lambda_{1/2} &= \frac{a+d}{2} \pm \sqrt{\frac{(a+d)^2}{4} - ad + b^2} \\ &= \frac{a+d}{2} \pm \sqrt{\frac{a^2+d^2}{4} + \frac{ad}{2} - ad + b^2} \\ &= \frac{a+d}{2} \pm \sqrt{\frac{(a-d)}{4} + b^2}.\end{aligned}$$

The expression in the square root is always non-negative, so that $\lambda_{1/2}$ are real numbers.

As an example of a non-symmetric matrix with imaginary eigenvalues, we have seen in Example 6.21 that the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has the eigenvalues $\lambda_{1/2} = \pm i$.

Example 7.38. The matrix $A = \begin{pmatrix} 5 & -4 & 2 \\ -4 & 5 & 2 \\ 2 & 2 & -1 \end{pmatrix}$ is symmetric. We compute an orthonormal eigenbasis by first computing the eigenvalues:

$$\chi_A(t) = -t^3 + 9t^2 + 9t - 81.$$

The eigenvalues and an eigenvector for them are as follows:

- $\lambda_1 = 9$, $v_1 = (-1, 1, 0)$,
- $\lambda_2 = 3$, $v_2 = (1, 1, 1)$,
- $\lambda_3 = -3$, $v_3 = (-1, -1, 2)$.

These three vectors are orthogonal; this is seen by direct computation. Alternatively, since the eigenvalues are all distinct, they are automatically orthogonal (Exercise 7.12). They are however not normal, dividing by their norm gives an orthonormal eigenbasis:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

7.5 Affine subspaces

Definition 7.39. Let V be a vector space. An *affine subspace* of V is a subset of the form

$$v_0 + W := \{v_0 + w \ |w \in W\}$$

for an appropriate vector $v \in V$ and a sub*space* $W \subset V$.

In other words, an affine subspace is obtained by translating a subspace (i.e., a sub-vector space) by a certain vector. For example, any line or a plane in $\mathbf{R}^3$ that is not necessarily passing through the origin is an affine subspace. A key example of an affine subspace is the solution set of a (not necessarily homogeneous) linear system

$$Ax = b.$$

Indeed, by Theorem 4.36 its solution set is precisely an affine subspace. See also the illustration in Remark 4.37.

Lemma 7.40. Let $X = v + W \subset \mathbf{R}^n$ be an affine subspace. If $X = v' + W'$ for any vector $v' \in \mathbf{R}^n$ and a subspace $W' \subset \mathbf{R}^n$, then the following holds:

- $W = W'$ and
- $v - v' \in W$.

In other words, the sub-vector space W is uniquely determined by X.

Proof. If $v + W = v' + W'$, then $v - v' \in W$. This implies

$$W' = -v' + v' + W' = -v' + X = \underbrace{-v' + v}_{\in W} + W = W.$$

Here we have used that for a subspace $A \subset \mathbf{R}^n$ (such as $A = W$), and an element $a \in A$, we have $a + A = A$. □

We can therefore define the dimension of an affine subspace as $\dim X = \dim W$, if $X = v + W$ as above.

Definition 7.41. Let X, X' be two affine subspaces. Let $W, W' \subset \mathbf{R}^n$ be the associated sub-vector spaces, as per Lemma 7.40.

- We say X *intersects* X' if $X \cap X' \neq \varnothing$.
- We say X is *parallel* to X' if $W \subset W'$ or if $W' \subset W$.
- We say that X is *skew* to X' if $W \cap W' = \{0\}$ and if $X \cap X' = \varnothing$.

Example 7.42. We examine the relative position of the lines

$$\begin{aligned} X &= (1,-3,5) + L(1,-1,2) = \{(1+t, -3-t, 5+2t) \mid s \in \mathbf{R}\} \\ X' &= (4,-3,6) + L(-1,1,2) = \{(4-t, -3+t, 6+2t) \mid t \in \mathbf{R}\}. \end{aligned}$$

The two subspaces W and W' are spanned by $(1,-1,2)$ and $(-1,1,2)$, respectively. These two vectors are linearly independent, so that the lines are not parallel. We determine whether they have an intersection point by solving the system

$$(1+s, -3-s, 5+2s) = (4-t, -3+t, 6+2t).$$

Considering the first two equations gives $s+t=3$ and $s+t=0$, which has no solution. Thus $X \cap X' = \varnothing$, which means that the lines are skew.

Definition 7.43. For two affine subspaces $X, X' \subset \mathbf{R}^n$ we say that two points $x \in X$, $x' \in X'$ *realize the minimal distance* of X and X' if

$$d(x, x') \leqslant d(x_1, x_1')$$

for any two points $x_1 \in X, x_1' \in X'$. In this event, we also write $d(X, X') := d(x, x')$ for that minimal distance.

Proposition 7.44. Let $X = v_0 + W$ be an affine subspace of a Euclidean space V. There is a unique vector $v \in V$ characterized by the following equivalent properties:
(1) v is an element of $X \cap W^\perp$,

(2) v realizes the minimal distance of the origin to W.
This vector v is given by

$$v = p_{W^\perp}(v_0) = v_0 - p_W(v_0),$$

i.e., the projection of v_0 onto the orthogonal complement $W^\perp$.

Proof. We first prove that $X \cap W^\perp$ contains v as defined above. Indeed, by Theorem 7.24, we can write

$$v_0 = w + v$$

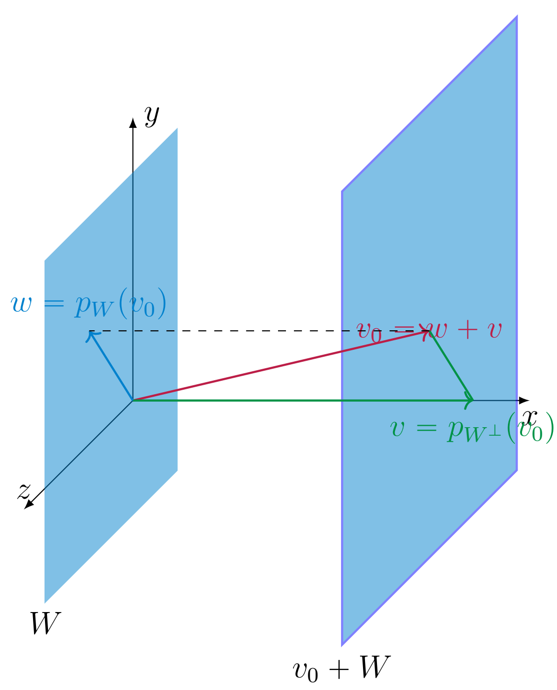
with uniquely determined $w = p_W(v_0) \in W$ and $v = p_{W^\perp}(v) \in W^\perp$. This means that $v = v_0 - w \in X \cap W^\perp$.

We now prove that $X \cap W^\perp$ consists only of that vector v. If another vector $v' \in W^\perp \cap X$, then $v' = v_0 + \tilde{w}$ for $\tilde{w} \in W$, so that $v - v' = w - \tilde{w} \in W \cap W^\perp = \{0\}$, so that $v = v'$.

We prove that this vector v realizes the minimal distance to the origin. To this end, let $x \in X$ be any vector. We need to prove $\|x\| \geqslant \|w\|$. Then $w := v - x \in W$. We can then compute

$$\begin{aligned} \|x\| &= \|v + w\| \\ &= \sqrt{\langle v + w, v + w \rangle} \\ &= \sqrt{\langle v, v \rangle + 2 \underbrace{\langle v, w \rangle}_{=0} + \langle w, w \rangle} \quad \text{by bilinearity and symmetry} \\ &= \sqrt{\|v\|^2 + \|w\|^2} \\ &\geqslant \|v\|. \end{aligned}$$

Here is a picture of the proof idea:



We finally show that a vector $x \in X$ with minimal distance to the origin agrees with v:

$$\begin{aligned}
\|x\|^2 &= \|\underbrace{x-v}_{=:w\in W}+v\|^2 \\
&= \langle w+v, w+v\rangle \\
&= \|w\|^2 + 2\underbrace{\langle w,v\rangle}_{=0} + \|v\|^2 && \text{(bilinearity)} \\
&= \|w\|^2 + \|v\|^2. && \text{since } v \in W^\perp
\end{aligned}$$

Since $\|x\| = \|v\|$, this implies $\|w\|^2 = 0$, i.e., $w = 0$, i.e., $x = v$. □

Definition 7.45. A *hyperplane* in $\mathbf{R}^n$ is an affine subspace H of dimension $n-1$, i.e., an affine subspace of the form

$$H = v_0 + W$$

where W is a subspace with $\dim W = n-1$.

For example, a line is a hyperplane in $\mathbf{R}^2$, and a plane is a hyperplane in $\mathbf{R}^3$.

Proposition 7.46. Let $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbf{R}^n$ be a *non-zero* (column) vector, and let $b \in \mathbf{R}$. Then the subset

$$H := \{x \in \mathbf{R}^n \mid \langle x, a\rangle = b\}$$

is a hyperplane. Its distance to the origin is given by

$$d(0,H) = \frac{|b|}{\|a\|}.$$

Proof. We show that H is a hyperplane. Indeed, the equation $\langle x, a\rangle = b$, which can be rewritten as

$$a_1x_1 + \cdots + a_nx_n = b$$

is a (non-homogeneous) linear system and the matrix $\begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}$ has rank 1, since the vector is nonzero. Therefore H has dimension $n-1$.

Let $W := \{x \in \mathbf{R}^n \mid \langle x, a\rangle = 0\}$ be the associated subspace. Then $H = v + W$ for some $v \in \mathbf{R}^n$, according to Theorem 4.36. Thus, $a \in W^\perp$. If we set $\lambda := \frac{b}{\|a\|^2}$, we have $\lambda a \in H$:

$$\left\langle \frac{b}{\|a\|^2}a, a\right\rangle = \frac{b}{\|a\|^2}\langle a, a\rangle = b.$$

Therefore $\lambda a \in H \cap W^\perp$. Thus, by Proposition 7.44, λa is the closest vector (in H) to the origin, and we have

$$d(0, H) = \|\lambda a\| = \frac{|b|}{\|a\|}.$$

Above we saw that an equation of the form

$$\langle x, a \rangle = b$$

for fixed $a \neq 0$ and $b \in \mathbf{R}$ determines a hyperplane. Here is a converse to this statement.

Proposition 7.47. (*Hesse normal form* of a hyperplane) Let $H = v_0 + W \subset \mathbf{R}^n$ be a hyperplane, and let $d = d(0, H)$ be its distance to the origin. Then there is a unique vector $a \in \mathbf{R}^n$ such that
(1) $\|a\| = 1$,

(2) $a \in H^\perp$,

(3) $H = \{x \in \mathbf{R}^n \mid \langle x, a \rangle = d\}$.
This vector can be computed as

$$a = \frac{v}{\|v\|},$$

where v is the unique element in $H \cap W^\perp$ or (equivalently) the point in H that is closest to the origin.

The equation $\langle x, a \rangle = d$ (which is a linear equation in the unknowns $x_1, \dots, x_n$) is called the *Cartesian equation* of the hyperplane.

Example 7.48. We continue the example in Example 7.22:

$$\begin{aligned} W &= L\left(\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right) \\ W^\perp &= L\left(\begin{pmatrix} 4 \\ -4 \\ 1 \end{pmatrix}\right), \end{aligned}$$

and consider the hyperplane $H = \begin{pmatrix} 11 \\ 11 \\ 11 \end{pmatrix} + W$. We compute $H \cap W^\perp$, which by Proposition 7.44 requires to find $w \in W$ and $v \in W^\perp$

such that

$$v_0 = \begin{pmatrix} 11 \\ 11 \\ 11 \end{pmatrix} = w + v$$
$$= a\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + b\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 4 \\ -4 \\ 1 \end{pmatrix}.$$

We compute the inverse of $A = \begin{pmatrix} 1 & 1 & 4 \\ 2 & 1 & -4 \\ 4 & 0 & 1 \end{pmatrix}$ using Theorem 4.80 (alternatively, one can also use the adjugate matrix, as in Theorem 5.13). The result is

$$A^{-1} = \frac{1}{33}\begin{pmatrix} -1 & 1 & 8 \\ 18 & 15 & -12 \\ 4 & -4 & 1 \end{pmatrix}.$$

According to Theorem 4.68, the above system therefore has a unique solution, given by

$$A^{-1}v_0 = \frac{1}{3}\begin{pmatrix} 8 \\ 21 \\ 1 \end{pmatrix}.$$

Thus, $c = \frac{1}{3}$ above, so that

$$v = \frac{1}{3}\begin{pmatrix} 4 \\ -4 \\ 1 \end{pmatrix}.$$

According to Proposition 7.44, this is the closest vector in H to the origin, and the distance of H to the origin is given by

$$d = \|v\| = \sqrt{\frac{33}{9}} = \sqrt{\frac{11}{3}}.$$

In addition,

$$a = \sqrt{\frac{11}{27}}\begin{pmatrix} 4 \\ -4 \\ 1 \end{pmatrix}.$$

7.5.1 Lower dimensional affine subspaces

This representation of hyperplanes can also be used to understand the geometry of subspaces of smaller dimension. For simplicity, we discuss this in the special case of lines in $\mathbf{R}^3$. A line $L \subset \mathbf{R}^3$ can be described in two ways:

(1) L can be described as an affine subspace

$$L = v + L(w) \tag{7.49}$$

for appropriate vectors $v, w \in \mathbf{R}^3$. I.e., the points in L are of the form $v + \lambda w$ for $\lambda \in \mathbf{R}$. This can be spelled out for each of the three components:

$$x_k = v_k + \lambda w_k \text{ for } k = 1, 2, 3. \tag{7.50}$$

This system is referred to as the system of *vector equations* or *parametric equations*.

(2) L can also be described by a system of two equations

$$\begin{aligned} a_1 x_1 + a_2 x_2 + a_3 x_3 &= b \\ a'_1 x_1 + a'_2 x_2 + a'_3 x_3 &= b'. \end{aligned} \tag{7.51}$$

This system is referred to as the system of *cartesian equations* of L. If we write $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ etc., it can be rewritten more compactly as

$$\begin{aligned} \langle x, a \rangle &= b \\ \langle x, a' \rangle &= b'. \end{aligned}$$

Each of these two equations describes a hyperplane in $\mathbf{R}^3$, i.e., a plane, and the line is the intersection of these planes.

One can pass from (1) to (2) by eliminating λ in (7.50). Conversely, in order to present L as an affine subspace, i.e., in the form

$$L = v + L(w),$$

we solve the above linear system.

Example 7.52. The following equations

$$\begin{aligned} x + y - 1 &= 0 \\ 3x + y - 2z - 1 &= 0 \end{aligned}$$

determine a line $L \subset \mathbf{R}^3$. We compute a representation $L = v + W$ by solving the system:

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 3 & 1 & -2 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & -2 & -2 & -2 \end{array}\right),$$

so the solutions are $y = 1 - z$, $x = 1 - y = z$, i.e.,

$$L = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + L\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right).$$

Example 7.53. The line

$$L = (1, 0, 1) + L(2, 1, -1),$$

is described by the vector equations

$$\begin{aligned} x_1 &= 1 + 2\lambda \\ x_2 &= \lambda \\ x_3 &= 1 - \lambda. \end{aligned}$$

The cartesian equations can be determined by observing that $\lambda = x_2$, so that the other two equations read

$$\begin{aligned} x_1 &= 1 + 2x_2 \\ x_3 &= 1 - x_2 \end{aligned}$$

which can be rewritten as

$$\begin{aligned} x_1 - 2x_2 &= 1 \\ x_2 + x_3 &= 1 \end{aligned}$$

or, yet equivalently

$$\begin{aligned} \left\langle x, \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix} \right\rangle &= 1 \\ \left\langle x, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\rangle &= 1. \end{aligned}$$

The planes P containing the given line L, i.e. such that $L \subset P$ can be characterized by the equation

$$\langle x, \lambda a + \lambda' a' \rangle = \lambda b + \lambda' b',$$

where $\lambda, \lambda' \in \mathbf{R}$ are arbitrary such that $\lambda a + \lambda' a' \neq 0$. Indeed, this equation does describe a (hyper)plane, and if $x \in L$, then it satisfies this latter equation.

Example 7.54. The line defined by the equations

$$L : x_2 = 0, x_3 = 1$$

can be written as $\left\langle x, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle = 0$ and $\left\langle x, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle = 1$. (It can also be written as $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + L(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix})$.) Thus, the planes P containing L are all of the form

$$\left\langle x, \begin{pmatrix} 0 \\ \lambda \\ \lambda' \end{pmatrix} \right\rangle = \lambda',$$

for arbitrary $\lambda, \lambda' \in \mathbf{R}$. Note that the vectors $\begin{pmatrix} 0 \\ \lambda \\ \lambda' \end{pmatrix}$ are precisely the vectors orthogonal to the vector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Given another line $L' = v' + L(w')$, this conveniently allows to determine the plane P that is parallel to L'. The line L' is parallel to P exactly if

$$w' \perp \lambda a + \lambda' a'$$

for appropriate $\lambda, \lambda' \in \mathbf{R}$.

Example 7.55. Continueing the example above, let

$$L' : z = 2, x = y.$$

It is given by $L' = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} + L(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix})$. We solve the equation

$$w' = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \perp \begin{pmatrix} 0 \\ \lambda \\ \lambda' \end{pmatrix},$$

it gives $\lambda = 0$, and $\lambda' \neq 0$ is arbitrary. Thus, for any λ', the plane defined by the equation

$$\left\langle x, \begin{pmatrix} 0 \\ 0 \\ \lambda' \end{pmatrix} \right\rangle = \lambda'$$

is parallel to L' and contains L. This gives the equation

$$\left\langle x, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle = 1$$

or, more concretely, $x_3 = 1$.

7.6 Distance between two affine subspaces

Theorem 7.56. Let $X = v + W$, $X' = v' + W'$ be two affine subspaces. Let us write $d := v - v'$ and $Z := W + W'$ (Definition 3.34). Let

$$m := p_{Z^\perp}(d) = d - p_Z(d)$$

be the orthogonal projection of d onto $Z^\perp$ (Corollary 7.33).

For two points $x \in X$ and $x' \in X'$ the following are equivalent:

(1) $x - x' = m$.

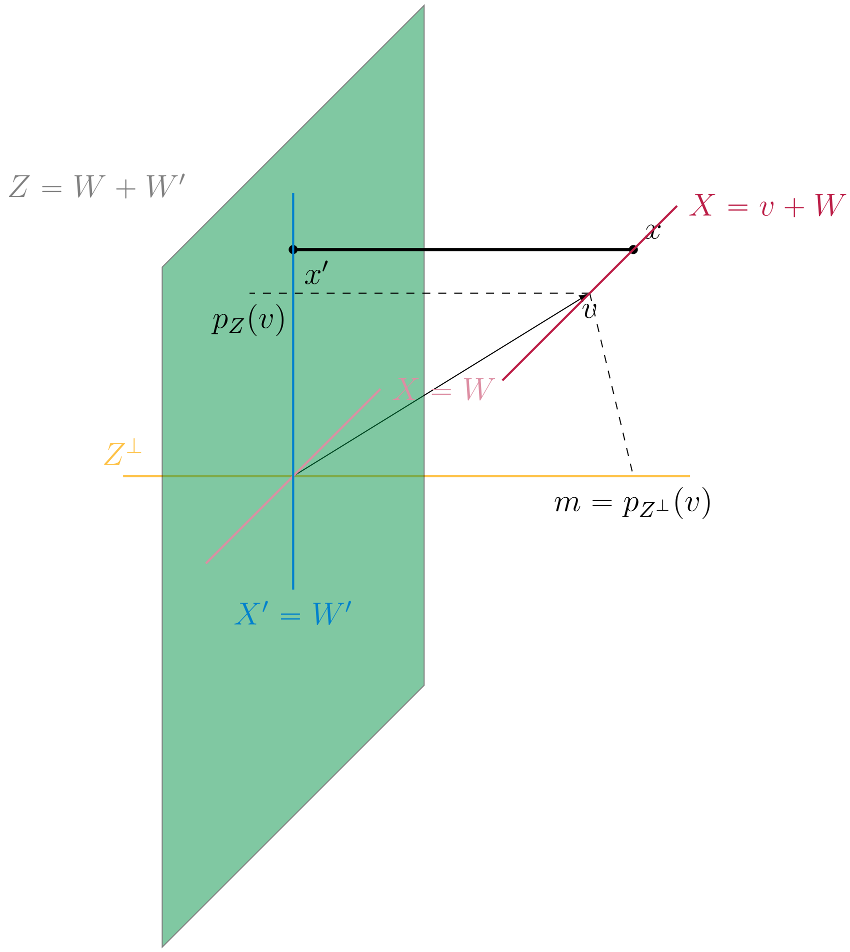
(2) $d(x, x') = \|m\|$.

(3) x and x' realize the minimal distance of X and X', i.e., $d(x, x') = d(X, X')$.

(4) The vector $x - x'$ is orthogonal to W and to W' (i.e., $x - x'$ is orthogonal to any $w \in W$, $w' \in W'$).

In particular, X intersects X' if and only if $d \in Z$.

Proof. Here is a picture of the geometric ideas in the proof. For simplicity of the picture, we choose $v' = 0$, so that $X' = W'$ and $d = v$.



(1) ⇒ (2) is obvious since $d(x, x') = \|x - x'\|$.

We next prove the equivalence (3) ⇔ (2). We write a point $x \in X$ as $x = v + w$ with an arbitrary vector $w \in W$. Likewise, $x' = v' + w'$. We then have

$$d(x, x') = \|x - x'\| = \|v - v' + w - w'\| = \|d + w - w'\|.$$

The vector $w - w'$ is an arbitrary vector in the sum $Z = W + W'$ (notice that for any $w' \in W'$, also $-w' \in W'$).

Therefore, we are seeking the point $z \in Z = W + W'$ such that $\|d + z\|$ is minimal. This is just the distance of the affine subspace $d + Z$ to the origin. According to Proposition 7.44, this distance is given by $\|m\| = \|p_{Z^\perp}(d)\| = \|d - p_Z(d)\|$, and m is the unique vector in Z realizing that minimal distance. This shows the equivalence of (3) and (2).

(3) $\Rightarrow$ (4): let $x \in X$ and $x' \in X'$ be two points realizing that minimal distance: $d(x, x') = \|m\|$. In particular, this means that $x' \in X'$ is the point realizing the minimal distance to x. Again by Proposition 7.44, $x' - x$ is therefore orthogonal to W'. Switching the role of X and X' we obtain similarly that $x - x'$ is orthogonal to W.

(4) $\Rightarrow$ (1): Our assumption means that

$$x - x' \in W^\perp \cap W'^\perp = (W + W')^\perp = Z^\perp.$$

To see the latter equality note that some vector is orthogonal to $W + W'$ precisely if it is orthogonal to W and to W', by the bilinearity of $\langle -, - \rangle$. We use this remark as follows: from

$$x - x' = v + w - v' - w'$$

we get

$$d = v - v' = \underbrace{x - x'}_{\in Z^\perp} + \underbrace{w' - w}_{\in Z}.$$

By the unicity of the representation of d as a sum of a vector in $Z^\perp$ and one in Z, this means that $x - x' = p_{Z^\perp}(d) = m$. □

Example 7.57. We consider the two lines in $\mathbf{R}^3$

$$X = (2, -1, 3) + L(1, 1, -2) = v + W$$
$$X' = (-3, 0, 0) + L(0, 2, 4) = v' + W'.$$

The general vectors of X and X' are of the following form, for $a, b \in \mathbf{R}$.

$$\begin{aligned} x &= (2, -1, 3) + a(1, 1, -2) &&= (2 + a, -1 + a, 3 - 2a) \\ x' &= (-3, 0, 0) + b(0, 2, 4) &&= (-3, 2b, 4b) \\ x - x' & &&= (5 + a, -1 + a - 2b, 3 - 2a - 4b) \end{aligned}$$

We compute the minimal distance of X and X' by considering the condition $x - x' \perp (1, 1, -2)$ and $x - x' \perp (0, 2, 4)$. This gives the following homogeneous linear system

$$\begin{aligned} 0 &= (5 + a) + (-1 + a - 2b) - 2(3 - 2a - 4b) &&= -2 + 6a + 6b \\ 0 &= 2(-1 + a - 2b) + 4(3 - 2a - 4b) &&= 10 - 6a - 20b. \end{aligned}$$

This can be solved to $b = \frac{4}{7}$ and $a = -\frac{5}{21}$. The points x and x' and their distance is then readily computed.

7.7 Summary of some computational tasks

(1) Given a subspace $U \subset \mathbf{R}^n$, compute an orthonormal basis: compute a basis, then orthonormalize the basis using the Gram–Schmidt algorithm (Proposition 7.31).

See Exercise 7.17 for a concrete example.

(2) Given some point $x \in \mathbf{R}^n$, compute its orthogonal projection to a sub-vector space $U \subset \mathbf{R}^n$: compute an orthonormal basis $u_1, \dots, u_k$ of U (cf. (1)), then

$$p_U(x) = \sum_{i=1}^{k} \langle x, u_i \rangle u_i.$$

See Exercise 7.2 for a concrete example. One may also compute $U^\perp$ (cf. (3)), and then compute $p_{U^\perp}(x)$, and then $p_U(x) = x - p_{U^\perp}(x)$. This approach may in practice be easier especially if $\dim U > \dim U^\perp$.

(3) Given a subspace $U \subset \mathbf{R}^n$, compute its orthogonal complement: compute a basis, say $u_1, \dots, u_k$. Then solve the homogeneous linear system given by the equations $\langle x, u_1 \rangle = 0, \dots, \langle x, u_k \rangle = 0$. I.e., form the matrix

$$\begin{pmatrix} u_1 \\ \dots \\ u_k \end{pmatrix}$$

(whose rows are the basis vectors u_1 etc.), and compute its kernel.

See Exercise 7.17 for a concrete example.

(4) Given a line L in cartesian equations (as in (7.51)), convert it into a vector equation as in (7.49): pick two (distinct) points p, q satisfying the cartesian equations. Then $L = p + L(p - q)$.

See Exercise 7.18 or Exercise 7.19 for a concrete example. Choosing different points instead of p, q will give different vector equations; but if $L = p' + L(p' - q')$, the two vectors $p - q$ and

$p' - q'$ will necessarily be (non-zero) multiples of each other (by Lemma 7.40).

(5) Given a line $L = v + L(w)$ (i.e., in vector equations), convert it into cartesian equations: express one of the three coordinates in terms of the free parameter, insert it into the others.

See Example 7.53 for a concrete example.

(6) Given three points $p, q, r \in \mathbf{R}^3$ (not lying on a line), describe the plane P containing p, q, r in cartesian form: We have $P = p + L(q - p, r - p)$. Similarly to (4), different choices of p, q, r may give different results (but $q-p, r-p$ need to span the same subspace as $q' - p', r' - p'$).

See Exercise 7.19 for a concrete example.

(7) Given a line L and a point r (with $r \notin L$), find the plane P containing L and r: choose two distinct points $p, q \in L$, then apply (6).

(8) Given a plane $P = \{x | \langle x, a \rangle = d\}$, a line $L = v + L(w) \subset P$ and a point $p \in P$, find the line $M \subset P$ that contains p and is orthogonal to L: $M = p + L(w')$, with w' being orthogonal to a and to w, i.e., $w' \in (L(a, w))^{\perp}$ (which can be computed as in (3)).

See Exercise 7.22 for a concrete example.

(9) Given a plane $P = v + L(w_1, w_2)$ in cartesian form, compute it as $P = \{x \in \mathbf{R}^3 | \langle x, a \rangle = d\}$: compute the orthogonal complement of $W = L(w_1, w_2)$. Pick a non-zero vector $a \in W^{\perp}$. Then compute $d = \langle x, v \rangle$.

See Exercise 7.19 for a concrete example. Note: a presentation of P as above is not unique, but another presentation as $P = \{x \in \mathbf{R}^3 | \langle x, a' \rangle = d'\}$ will be such that $a = \lambda a'$ for some $\lambda \in \mathbf{R}, \lambda \neq 0$, and $d = \lambda d'$.

(10)[redacted]heck whether two lines $L = v + W = v + L(w)$ and $L = v' + W' = v' + L(w')$ are parallel: this is the case precisely if w and w' are linearly dependent. (If the lines are given in cartesian form, apply (4) first.)

See Exercise 7.8 and Exercise 7.18 for concrete examples.

(11)[redacted]heck whether two lines $L = v + W = v + L(w)$ and $L = v' + W' = v' + L(w')$ are skew: this is the case precisely if a) w

and w' are linearly independent and b) $L \cap L' = \varnothing$. To check the condition b), it is convenient to express the lines in cartesian equations first (cf. (5)).

See Exercise 7.18 for a concrete example.

(12) Check whether a line $L = v + W = v + L(w)$ is parallel to a plane $P = v' + W' = v' + L(w'_1, w'_2)$: this is the case precisely if $W \subset W'$ or, equivalently, if w is a linear combination of w'_1 and w'_2. If the plane P is instead given as $P = \{x | \langle x, a \rangle = d\}$, L is parallel to P precisely if $\langle w, a \rangle = 0$, i.e., if w (the direction vector of L) is orthogonal to the plane W' underlying P.

See Exercise 7.18(3) and Exercise 7.23 for a concrete example.

(13) Check whether a line $L = v + W = v + L(w)$ is orthogonal to a plane $P = \{x | \langle x, a \rangle = d\}$: this is the case precisely if w and a are linearly dependent. If the plane is given rather as $P = v' + W' = v' + L(w'_1, w'_2)$, it is the case precisely if $\langle w, w'_1 \rangle = 0$ and $\langle w, w'_2 \rangle = 0$.

(14) Compute the distance of two affine subspaces $X = v + W$ and $X' = v' + W'$ (this includes the case when, say, X is a point, in which case $W = \{0\}$, and $X = v$):

- apply Theorem 7.56 via part (1): compute $Z = W + W'$, $d := v - v'$, and compute the orthogonal projection $m = p_{Z^\perp}(d)$, or equivalently $m = d - p_Z(d)$ (cf. (2)). Let x be a general point in X, x' a general point in X', and solve the linear system $x - x' = m$, – or –,
- apply Theorem 7.56 via part (4): again let x be a general point in X, x' a general point in X', compute $x - x'$ and solve the homogeneous linear system given by $\langle x, w \rangle = 0$ and $\langle x, w' \rangle = 0$, where w runs through a set of vectors spanning W, and w' runs through a set of vectors spanning W'.

See Exercise 7.9 or Exercise 7.23 for a concrete example.

7.8 Exercises

Exercise 7.1. Let $V = P_{\leqslant 2} = \{at^2 + bt + c \mid a, b, c \in \mathbf{R}\}$ be the vector space of (real) polynomials of degree $\leqslant 2$. We consider the scalar product in Example 7.16(3), i.e.,

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx.$$

- Let $e_1 = 1$, $e_2 = t$ and $e_3 = t^2$. (These vectors form a basis of $P_{\leqslant 2}$.) Compute $\langle e_i, e_j \rangle$ for $1 \leqslant i, j \leqslant 3$.
- Apply the Gram–Schmidt orthogonalization procedure to this basis.

Exercise 7.2. (Solution at p. 269) Consider the subspace $U \subset \mathbf{R}^3$ given by the solutions of the homogeneous linear system

$$x - y + 3z = 0.$$

(1) Find a basis of U.

(2) Compute a basis of $U^\perp$. What is $\dim U^\perp$?

(3) Consider $t = (0, 1, 5)$. Find its orthogonal projection onto U (recall from Corollary 7.33 that $t = t_U + t_\perp$ with uniquely determined vectors $t_U \in U$ and $t_\perp \in U^\perp$. The orthogonal projection of t onto U is then the vector t_U.)

Exercise 7.3. Consider the subspace $W \subset \mathbf{R}^4$ given by the equations

$$\begin{aligned} x - t &= 0 \\ y + z - t &= 0 \end{aligned}$$

(where x, y, z, t are the coordinates of $\mathbf{R}^4$).

(1) Compute a basis of W and of $W^\perp$.

(2) Compute the orthogonal projection of $t = (1, 5, 1, 6)$ onto W.

Exercise 7.4. (Solution at p. 271) Consider the subspace $U \subset \mathbf{R}^3$ given by the equations

$$\begin{aligned} x &= 0 \\ x + y + z &= 0 \end{aligned}$$

(where x, y, z are the coordinates of $\mathbf{R}^3$).

(1) Compute a basis of U and of $U^\perp$.

(2) Compute the orthogonal projection of $t = (5, 1, 3)$ onto U.

Exercise 7.5. (Solution at p. 272) Compute the orthogonal complement of $T = L((1, 0, -3))$.

Exercise 7.6. (Solution at p. 272) Is there a subspace $U \subset \mathbf{R}^3$ such that

(1) the orthogonal projection of $t = (1, 1, 0)$ onto U is given by $(1, 5, 6)$?

(2) the orthogonal projection of $t = (2, 0, 1)$ onto U is given by $(1, 1, 1)$?

Exercise 7.7. (Solution at p. 272) Let $L = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + L(\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix})$. Compute the closest point of L to the origin, and its distance to the origin.

Exercise 7.8. (Solution at p. 273) Consider the two lines $L : x = 1 + t, y = t, z = 2 + t, t \in \mathbf{R}$ and $L' : x - 3 = y - 1 = z - 3$. Are they parallel? Compute the distance between L and L'.

Exercise 7.9. (Solution at p. 273) Are the lines

$$L : x = y - 1 = -z \text{ and } L' : x - 2 = -y = \frac{z}{2}$$

identical, parallel, or skew? Compute their distance.

Exercise 7.10. (Solution at p. 274) Let P be the plane given by the equation

$$4x + 5y + 10z - 20 = 0.$$

Let L be the line given by the equations $x = 0, y = 5 - z$.

(1) Sketch P and L.

(2) Compute the orthogonal complement of the underlying vector space W of P.

(3) Compute the point of P that is closest to the origin and its distance to the origin.

(4) Are P and L parallel?

Exercise 7.11. (Solution at p. 276) Which of the following matrices is orthogonally diagonalizable? If so, find a orthonormal eigenbasis of $\mathbf{R}^2$.

(1) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

(2) $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$

(3) $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Exercise 7.12. (Solution at p. 277) Let A be a symmetric matrix and $\lambda \neq \mu$ two *distinct* eigenvalues of A, with eigenvectors v and w, respectively. Then $v \perp w$, i.e., eigenvectors of *distinct* eigenvalues are orthogonal.

Exercise 7.13. (Solution at p. 277) Let $P \subset \mathbf{R}^4$ by the hyperplane given by

$$2x_1 + x_3 - x_4 = 4,$$

where $(x_1, \dots, x_4)$ are the coordinates of $\mathbf{R}^4$. For a parameter $t \in \mathbf{R}$, let L_t be the line

$$L_t = (1, 0, 0, -2t) + L(t, 1, 0, -1).$$

- For which $t \in \mathbf{R}$ is L_t parallel to P?
- Let now $t = -\frac{1}{2}$ and consider the line $L = L_{-\frac{1}{2}}$. Determine the pair(s) of points (p, l) such that $p \in P$ and $l \in L$ such that their distance is minimal.

Exercise 7.14. Let $L \subset \mathbf{R}^3$ be the line defined be the system

$$\begin{aligned} x + z &= 1 \\ y + z &= -2. \end{aligned}$$

Let L' be the line in $\mathbf{R}^3$ passing through the points $(0, 0, 1)$ and $(0, 1, 1)$.

- Present L as $L = v + W$ for a subspace $W \subset \mathbf{R}^3$. Do the same for L'.
- Are L and L' a) identical, b) parallel or c) skew?

- Compute the Cartseian equation (i.e., in the form $ax+by+cz = d$, for appropriate values of $a, \dots, d$) of the plane $P \subset \mathbf{R}^3$ that contains L and is parallel to L'.
- Let $l = (2, -1, -1) \in L$. Compute a point $l' \in L'$ such that the line passing through l and l' is parallel to the plane given by the equation $x + z - 1 = 0$.

Exercise 7.15. Let $V = P_{\leqslant 2}$ be the vector space of polynomials of degree $\leqslant 2$. We write elements of V as $p(t) = a + bt + ct^2$, where $a, b, c \in \mathbf{R}$. Define

$$\langle p, q \rangle := \int_{-1}^{1} p(t)q(t)dt.$$

- Confirm that $\langle -, - \rangle$ is a scalar product on V.
- Compute an orthonormal basis of V.
- Consider the map

  $$f : V \to V, f(p) := t\frac{\partial p}{\partial t}$$

  (i.e., it maps a polynomial p to the product of the indeterminate t with the derivative of p with respect to the variable t). Confirm that this map is linear. Compute the matrix of f with respect to the standard basis $e_1 = 1$, $e_2 = t$ and $e_3 = t^2$. Is this basis an eigenbasis for f? Compute $\dim \ker f$ and $\dim \operatorname{im} f$.
- Does the map f have an orthonormal eigenbasis?

Exercise 7.16. Consider the subspace $U \subset \mathbf{R}^4$ given by the solutions of the equation

$$x_1 - x_2 + x_3 + 2x_4 = 0.$$

(As usual $x_1, \dots, x_4$ are the coordinates of $\mathbf{R}^4$.)

(1) Find a basis of U. What is $\dim U$?

(2) Compute an orthonormal basis of U.

(3) Compute the orthogonal projection of $v = (2, 3, 0, 0)$ and of $w = (2, 5, 3, 0)$ onto U.

(4) Compute $U^{\perp}$.

Exercise 7.17. (Solution at p. 279) In the Euclidean space $\mathbf{R}^4$, endowed with the standard scalar product, let U be the subspace spanned by the vectors $u_1 = (1, 2, 0, -1)$, $u_2 = (0, -4, 3, 4)$.

(1) Compute an orthogonal basis of U.

(2) Compute a basis of $U^\perp$.

(3) Compute the orthogonal projection of $v = (0, 5, 3, 4)$ onto U.

(4) Let $w = (2, -1, 0, 2)$. Decide whether there is a subspace $L \subset \mathbf{R}^4$ such that the orthogonal projection of w onto L is the vector $\ell = (1, 1, 2, 0)$.

Exercise 7.18. (Solution at p. 280) Consider the following two lines in $\mathbf{R}^3$, where x, y, z are the coordinates:

$$L : \begin{cases} x + y - 1 = 0 \\ 2x - z - 1 = 0 \end{cases} \qquad M : \begin{cases} x - 2y - 1 = 0 \\ y - z + 2 = 0 \end{cases}$$

(1) Determine whether L and M are the same line, parallel, or skew.

(2) Compute the cartesian equation of the plane that contains the line M and that is parallel to L. (Recall that a cartesian equation is of the form $\langle x, a \rangle = d$ for an appropriate vector a and an appropriate $d \in \mathbf{R}$.)

(3) Given the point $l = (0, 1, -1) \in L$ compute a point $m \in M$ such that the line passing through l and m is parallel to the plane defined by the equation $3x - z = 0$.

(4) Consider the family of planes $\pi_\alpha : z = \alpha$, for some parameter $\alpha \in \mathbf{R}$. Let $r_\alpha = L \cap \pi_\alpha$ and $s_\alpha = M \cap \pi_\alpha$. Let m_α be the midpoint of the segment with endpoints r_α and s_α. Verify that the points m_α are all lying on the same line. Moreover, determine the parametric equation of that line.

Exercise 7.19. (Solution at p. 281) Consider the points $p = (3, 1, 0)$, $q = (0, 1, 3)$ and $r = (-3, 0, -3) \in \mathbf{R}^3$. Let L be the line passing through p and q.

(1) Determine the parametric equation of L, i.e., express L in the form $L = v + W$, for an appropriate vector $v \in \mathbf{R}^3$ and a subspace $W \subset \mathbf{R}^3$.

(2) Verify that r does not lie on L. Give the plane P containing p, q, r both in vector and in cartesian form.

Exercise 7.20. (Solution at p. 282) Consider the line $L = (3, 1, 0) + L(1, 0, -1)$. Is there a plane containing L and the line M given by the systen $x + z = 2$, $x - 2y = 2$ (with x, y, z being the coordinates of $\mathbf{R}^3$)?

Exercise 7.21. (Solution at p. 282) Consider the line $L = (3, 1, 0) + L(1, 0, -1)$. Let $p = (-1, -1, -1)$. Describe all the points $q \in \mathbf{R}^3$ such that the line M passing through p and q intersects L orthogonally (i.e., intersects it, and does so orthogonally).

Exercise 7.22. (Solution at p. 283) Consider the plane $P = \{x \in \mathbf{R}^3, 3x - 4y + z = 2\}$ and the point $p = (0, 1, 6) \in P$, as well as the line $L = (0, 0, 2) + L(1, 1, 1) \subset P$, compute the line $M \subset P$ that is orthogonal to L and contains the point p.

Exercise 7.23. (Solution at p. 284) Consider the lines (in $\mathbf{R}^3$)

$$\begin{aligned} L_1 : 2x - y &= -3 \\ y + z &= -2 \end{aligned}$$

and

$$\begin{aligned} L_2 : x &= t \\ y &= 2 \\ z &= 4 - t, t \in \mathbf{R}. \end{aligned}$$

(1) Determine their relative position (skew, parallel etc.)
(2) Find the plane π parallel to L_1 and containing L_2.
(3) Compute the distance between π and L_1.
(4) Find points $p_1 \in L_1$, $p_2 \in L_2$ such that

$$\|p_1 - p_2\| = d(p_1, p_2) = d(L_1, L_2),$$

i.e., two points that realize the minimal distance between L_1 and L_2.

Exercise 7.24. (Solution at p. 286) We endow $\mathbf{R}^4$ with its usual scalar product. We consider the subspace $U \subset \mathbf{R}^4$ defined by the equations

$$\begin{cases} x_1 + x_3 = 0 \\ 2x_1 + x_2 - x_4 = 0 \end{cases}$$

(1) Compute an orthogonal basis of U.

(2) Given the vector $w_1 = (1, 1, -1, -1)$ find a vector w_2 that is orthogonal to w_1 and such that the vector space $W := L(w_1, w_2)$ satisfies $W = U^{\perp}$.

(3) Write down a system of linear equations in the unknowns x_1, x_2, x_3, x_4 whose solution set is the subspace $W = U^{\perp}$.

(4) Given the vector $v = (3, 1, -1, 1)$ find a vector $u \in U$ such that the vector $v + u$ has minimal norm.

Exercise 7.25. (Solution at p. 287) Consider the lines (contained in $\mathbf{R}^3$)

$$L : \begin{cases} x - 2y + 4 = 0 \\ y + z - 3 = 0 \end{cases} \qquad M : \begin{cases} x + 2z - 5 = 0 \\ x - 2y + 1 = 0 \end{cases}$$

(1) Verify that L and M are parallel and write down the cartesian equation of the plane containing L and M.

(2) Given the point $p = (0, 2, 1) \in L$ find the point $q \in M$ such that the line passing through p and q is orthogonal to L and to M.

(3) Write down the cartesian equation of the plane X containing the line L and passing through the point $r = (-1, 1, 0)$.

(4) Write down the parametric equation of the line N contained in the plane X, passing through $r = (-1, 1, 0)$ and orthogonal to the line L.

Exercise 7.26. (Solution at p. 289) Let $U \subset \mathbf{R}^4$ be the subspace defined by the equations

$$U : \begin{cases} x_1 - 2x_2 + x_4 = 0 \\ 3x_2 + x_3 - 2x_4 = 0 \\ 3x_1 + 2x_3 + t\,x_4 = 0 \end{cases}$$

(1) Find the value of t for which U has dimension 2. For this value of t, compute a basis of U.

(2) Apply the Gram–Schmidt method to the basis computed in part (a) in order to compute an orthogonal basis of U.

(3) For the value of t computed in part (a), compute a basis of $U^{\perp}$.

(4) Given $v = (3, 2, 2, -2) \in \mathbf{R}^4$ compute the cartesian equation of a subspace W of dimension 3 such that the orthogonal projection of v onto W is equal to the vector $w = (1, 2, 1, 1)$.

Exercise 7.27. (Solution at p. 290) We consider the following points (in $\mathbf{R}^3$) $A = (6, -1, -4)$, $B = (1, 1, -1)$ and the plane $X : 2x - y - 2z = 3$.

(1) Verify that $B \in X$. Let C be the orthogonal projection of A onto the plane X. Compute C.

(2) Compute the cartesian equation of the plane containing the triangle $\triangle ABC$.

(3) Compute the parametric equation of the line that a) passes through B, b) is contained in the plane X and c) is orthogonal to the line passing through A and B.

(4) Compute the value of the parameter t such that the line M_t : $\begin{cases} t\,x - y + 2 = 0 \\ x + z + 1 = 0 \end{cases}$ is parallel to the plane X.

Appendix A

Mathematical notation and terminology

Sets

Symbol	Reads	Explanation	Example
$\{\ \dots\}$	a set	The elements of the set are written inside the braces.	$\{1, 2, 3\}$ denotes the set consisting of the numbers 1, 2 and 3.
$\{\dots \mid \dots\}$	The set of all … satisfying the condition ….	This denotes the set consisting of all objects satisfiying a certain condition.	$\{$ all vegetables $V \mid$ I eat V regularly $\}$ consists of all the vegetables that I eat regularly.
$\in$	is an element of	If M is a set the expression $x \in M$ means that x is a member of M.	$\diamondsuit \in \{\diamondsuit, \heartsuit, \spadesuit, \clubsuit\}$
$\notin$	is not an element of	If M is a set the expression $x \in M$ means that x is a member of M.	$\diamondsuit \notin \{\heartsuit, \spadesuit, \clubsuit\}$

$f : X \to Y$	f from X to Y	A function f from a set X to another set Y.	$f : \{\text{Monday}, \dots, \text{Sunday}\} \to \{\text{true, false}\}$ is some function that assigns to any weekday either true or false. For example, f could indicate whether I go to school that day.
$\to$	to	The regular arrow is the symbol for a function.	
$\mapsto$	maps to	$x \mapsto y$ indicates that a particular element $x \in X$ is sent to (or "mapped to") the element $y \in Y$.	Sunday $\mapsto$ false
$X \times Y$	The *product* of two sets X and Y.	The product consists of pairs (x, y), where $x \in X$ and $y \in Y$.	$\{0, 1\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.
$X \subset Y$	*subset*	X is a subset of Y if every element of X is also an element of Y.	$\{1, 2\} \subset \{0, 1, 2\}$
$X \subsetneq Y$	*proper subset*	X is a proper subset of Y if $X \subset Y$ but $X \neq Y$	$\{1, 2\} \subsetneq \{0, 1, 2\}$
$X \cap Y$	*intersection* of X and Y	The intersection consists of those elements that are contained in X *and* in Y. $\{0, 1\} \cap \{-1, 0\} = \{0\}$	
$X \cup Y$	*union* of X and Y	The union consists of those elements that are contained in X *or* in Y. $\{0, 1\} \cap \{-1, 0\} = \{-1, 0, 1\}$	
$g \circ f$	*composition*	If $f : X \to Y$ and $g : Y \to Z$ are two functions, then $g \circ f : X \to Z$ is the function sending $x \in X$ to $g(f(x))$.	

Logic

Symbol	Reads	Explanation	Example
$\Rightarrow$	Implies	If A and B are two (mathematical) statements, then "$A \Rightarrow B$" means that if A holds then B also holds.	$x \geqslant 1 \Rightarrow x^2 \geqslant 1$
$\Leftrightarrow$	Equivalent	If A and B are two mathematical statements, then "$A \Leftrightarrow B$" is an abbreviation for $A \Rightarrow B$ *and* (at the same time) $B \Rightarrow A$.	$x \geqslant 0 \Leftrightarrow x + 1 \geqslant 1$
$:=$	is defined to be		$x := 2$ means that we define the variable x to take the value 2

Numbers and arithmetic

Symbol	Reads	Explanation	Example
$\mathbf{Z}$		The set of all integers.	$-34, -1, 0, 1, 2, 18, \cdots \in \mathbf{Z}$, $\frac{3}{4} \notin \mathbf{Z}$
$\mathbf{Q}$		The set of all rational numbers.	$\frac{-3}{16}, -3.3, -1, 0, 2.4, \frac{3}{4} \in \mathbf{Q}$, $\sqrt{3} \notin \mathbf{Q}$
$\mathbf{R}$		The set of all real numbers.	0, 1, -1, $\frac{1}{2}$, $\sqrt{3}$, π, $e \in \mathbf{R}$
$\sum_{e=1}^{n} a_e$	Sum	This is an abbreviation for the sum of the a_e, where e runs from 1 to n. (Here a_e can be any expression depending on e.) It can also be written as $a_1 + a_2 + \cdots + a_e$.	$\sum_{e=1}^{3} e^2 = 1^2 + 2^2 + 3^2 = 14$.

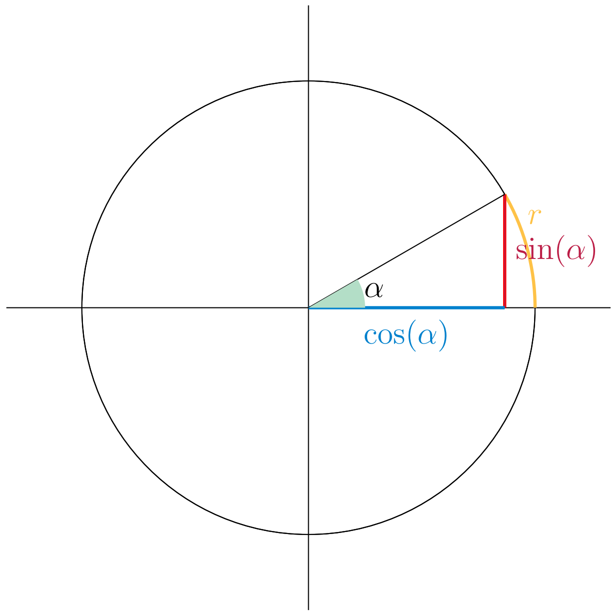
Appendix B

Trigonometric functions

Angles can be measured in degrees or in radians. These are converted as follows:

angle	radian
(in degree)	(no unit)
180°	π
90°	$\frac{\pi}{2}$
α	$\frac{\pi}{180}\alpha$
$\frac{180}{\pi}r$	r

Geometrically, given an angle α (between 0 and 360° as in the picture below), the radian is the length of the yellow circle segment as shown:



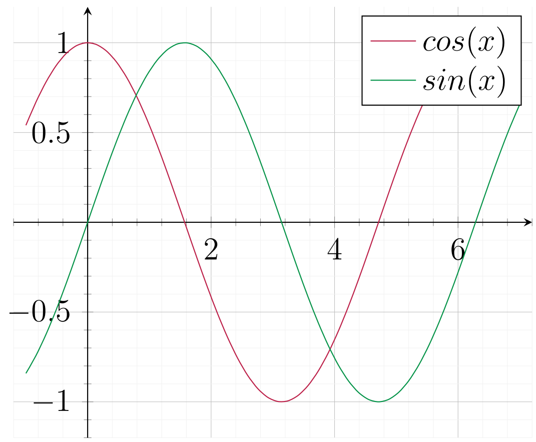
A rotation by a positive number is *counter-clockwise*; conversely negative numbers correspond to a *clockwise* rotation. For example, a rotation by $\frac{\pi}{2}$ is a counter-clockwise rotation by 90°. A rotation by $-\frac{\pi}{4}$ is a clockwise rotation by 45°.

Given any radian r, the ray that has an angle r between itself and the positive x-axis meets the circle with radius 1 and mid-point $(0, 0)$ in exactly one point p. The *trigonometric functions* sin and cos are defined to be the coordinates of that point:

$$p = (\cos(r), \sin(r)).$$

For example, we have the following values

r	0	$\pi/6$ (30°)	$\pi/4$ (45°)	$\pi/3$ (60°)	$\pi/2$ (90°)	...
$\sin(r)$	0	$\frac{1}{2}$	$\sqrt{2}$	$\frac{\sqrt{3}}{2}$	1	...
$\cos(r)$	1	$\frac{\sqrt{3}}{2}$	$\sqrt{2}$	$\frac{1}{2}$	0	...



cos(x)
sin(x)
1
0.5
−0.5
−1
2
4
6

Appendix C

Solutions of selected exercises

C.1 Complex numbers

Solution of Exercise 1.4: We have $\frac{i-1}{i+1} = \frac{(i-1)\overline{i+1}}{|i+1|^2} = \frac{(i-1)(1-i)}{2} = i$. Thus $z = i^3 = -i$ is the algebraic form. We have $|z| = 1$ and $\arg z = \frac{3}{2}\pi$, so

$$z = \cos(\frac{3}{2}\pi) + i\sin\frac{3}{2}\pi.$$

Solution of Exercise 1.5: In order to solve

$$z = 3i|z|\overline{z}$$

we would like to divide by z. This is only possible if $z \neq 0$, so we first consider the case $z = 0$. In this case both sides of the equation are equal to 0, so $z = 0$ is indeed a solution. Now, we consider $z \neq 0$ and divide the above equation by z and obtain

$$1 = 3i|z|\frac{\overline{z}}{z}.$$

There are different ways to solve this equation. One may put $z = a + ib$ and solve the resulting quadratic equation. For illustrational purposes, we rather consider the trigonometric form $z = r(\cos\alpha + i\sin\alpha)$. Then $\overline{z} = r(\cos\alpha - i\sin\alpha) = r(\cos(-\alpha) + i\sin(-\alpha))$, and

$$|z|\frac{\overline{z}}{z} = r\frac{r(\cos(-\alpha) + i\sin(-\alpha))}{r(\cos(\alpha) + i\sin(\alpha))}.$$

Note that $r = |z| \neq 0$, so we can cancel this in the right-hand fraction. We have

$$\frac{\cos(-\alpha) + i\sin(-\alpha)}{\cos(\alpha) + i\sin(\alpha)} = \cos(-2\alpha) + i\sin(-2\alpha).$$

We then obtain

$$\frac{1}{3i} = -\frac{1}{3}i = |z|\frac{\overline{z}}{z} = r(\cos(-2\alpha) + i\sin(-2\alpha)).$$

This implies $r = \frac{1}{3}$. Concerning the arguments, we have to be more careful: the above equation is equivalent to saying that

$$-2\alpha \equiv \frac{3}{2}\pi \mod 2\pi$$

cf. around (1.6). There are two solutions: $-2\alpha = \frac{3}{2}\pi$ or $-2\alpha = \frac{3}{2}\pi + 2\pi$. The former yields $\alpha = -\frac{3}{4}\pi$, the latter $\alpha = \frac{\pi}{4}$. (Of course, we can now add integer multiples of 2π to these values of α, so $\alpha = \frac{5}{4}\pi$ is another solution. However, this gives the same value for z.) The resulting solutions are

$$z = \frac{1}{3}(\cos(-\frac{3}{4}\pi) + i\sin(-\frac{3}{4}\pi))$$

and

$$z = \frac{1}{3}(\cos(\frac{1}{4}\pi) + i\sin(\frac{1}{4}\pi)).$$

To sum up, the above equation has three solutions, $z = 0$ and these two solutions.

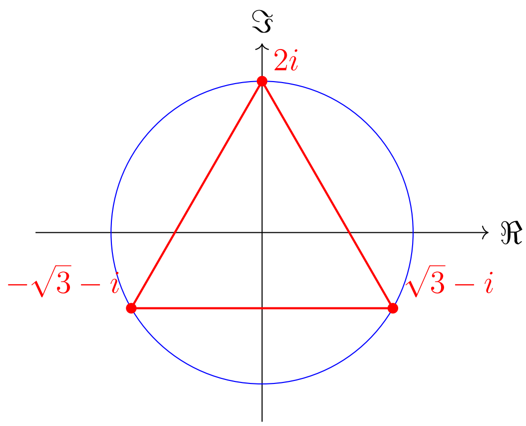
Solution of Exercise 1.6: There are three solutions, namely

$$z_0 = \sqrt{3} - i = 2(\cos(-\pi/6) + i\sin(-\pi/6)),$$

$$z_1 = -\sqrt{3} - i = 2(\cos(-5\pi/6) + i\sin(-5\pi/6)),$$

$$z_2 = 2i = 2(\cos(-3\pi/2) + i\sin(-3\pi/2)).$$

Here is a picture:



C.2 Systems of linear equations

Solution of Exercise 2.6: If $a \neq 0$ *or* $b \neq 0$, then the equation $ax + by = c$ has infinitely many solutions. Indeed, if, say $a \neq 0$, we can subtract by and divide by a, which gives $x = \frac{c-by}{a}$. Thus, for any $y \in \mathbf{R}$, the pair $(x = \frac{c-by}{a}, y)$ is a solution. A similar analysis works if $b \neq 0$. It remains to consider the case in which $a = 0$ and $b = 0$. In this case the solution set of the equation depends on c:

- If $c = 0$, then *any* pair (x, y) is a solution. Indeed: $0x + 0y = 0$ holds true then. Thus, if $a = b = c = 0$, there are infinitely many solutions.
- If $c \neq 0$, the equation $0x + 0y = c$ has no solution, since the left hand side is always 0, while the right hand side is nonzero. So, in the case $a = b = 0$ but $c \neq 0$, there is *no* solution.

Solution of Exercise 2.10: The matrix associated to the system is as follows, and we bring it to reduced row echelon form:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ -2 & -3 & 1 & 1 \\ 0 & 1 & -1 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} \underline{1} & 0 & +1 & -2 \\ 0 & \underline{1} & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \blacksquare$$

We have two leading ones (underlined), so the third unknown x_3 is a free variable and x_1 and x_2 are non-free, and we have $x_2 = 1 + x_3$ and $x_1 = -2 - x_3$. Thus, the solution set is

$$\{(-2 - x_3, 1 + x_3, x_3) \mid x_3 \in \mathbf{R}\}.$$

Solution of Exercise 2.12: We apply Method 2.32. The matrix associated to the system is

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 & 1 \\ 1 & -1 & 0 & 1 & -3 \\ 1 & -1 & 3 & -2 & 0 \end{array}\right).$$

We compute the reduced row-echelon form of that matrix using Gaussian eliminiation (Method 2.30): we subtract the first row from the third, which gives

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 & -1 \\ 1 & -1 & 3 & -2 & 0 \end{array}\right).$$

We then subtract the first row from the fourth:

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 2 & -2 & 2 \end{array}\right).$$

We add the second line to the third:

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 2 \end{array}\right).$$

We then add (-2) times the second line to the fourth (equivalently, subtract 2 times the second line from the fourth):

$$\left(\begin{array}{cccc|c} \underline{1} & -1 & 1 & 0 & -2 \\ 0 & 0 & \underline{1} & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This matrix is in row-echelon form, with the leading 1's being underlined above. We finally bring it into reduced row-echelon form by subtracting the second from the first line, which gives

$$\left(\begin{array}{cccc|c} \underline{1} & -1 & 0 & 1 & -3 \\ 0 & 0 & \underline{1} & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The matrix has no entry of the form $0 \ \ldots \ 0 \ 1$, so the system does have a solution. The first column of the matrix corresponds to the variable x_1 etc., so that the free variables are x_2 and x_4. We let $x_2 = \alpha$, $x_4 = \beta$, where α and β are arbitrary real numbers. The non-free variables x_1 and x_3 are uniquely determined by α and β. To compute them, we use the equations obtained by the matrix

$$\begin{aligned} x_3 - \beta &= 1 \\ x_1 - \alpha + \beta &= -3 \end{aligned}$$

which we solve as $x_3 = 1 + \beta$ and $x_1 = \alpha - \beta - 3$. Thus, the solution set is

$$\{(\alpha - \beta - 3, \alpha, 1 + \beta, \beta) \mid \alpha, \beta \in \mathbf{R}\}.$$

Solution of Exercise 2.14: Hint: we will apply Gaussian elimination, but it simplifies the calculations to do a certain change of rows first. (Why is that allowed?)

Solution of Exercise 2.17: Suppose $x_1 = 1 - t$, $x_2 = 2 + 3t$ and $x_3 = 4t$. We have to determine whether there is some $t \in \mathbf{R}$ such that for these choices of x_1, x_2, x_3, we have a solution of the given system, i.e., whether

$$\begin{aligned} x_1 + x_2 + x_3 &= (1 - t) + (2 + 3t) + 4t &&= 1 \\ x_1 - x_3 &= 1 - t - 4t &&= 0. \end{aligned}$$

Simplifying these equations gives the system

$$\begin{aligned} 6t + 3 &= 0 \\ 1 - 5t &= 0. \end{aligned}$$

This system has no solutions, so there is no $t \in \mathbf{R}$ such that the vector $(1 - t, 2 + 3t, 4t)$ is a solution to the original system.

Solution of Exercise 2.19: We substitute $x_1 = 1 + t$, $x_2 = t + q$ and $x_3 = -t + 2q + 1$ into the given equation and get the equation

$$3(1 + t) + 2(t + q) - (-t + 2q + 1) = 5.$$

This simplifies to

$$6t + 2 = 5$$

which has the solution $t = -\frac{1}{2}$. Since the variable q does not appear in that equation it is a free variable. Thus, for all $q \in \mathbf{R}$, the vector

$$(x_1 = 1 - \frac{1}{2}, x_2 = -\frac{1}{2} + q, x_3 = \frac{1}{2} + 2q + 1) = (\frac{1}{2}, -\frac{1}{2} + q, \frac{3}{2} + 2q)$$

satisfies the requested conditions. Note that these are infinitely many solutions.

Solution of Exercise 2.20: We have to find $a_0, \dots, a_3$, so these are the unknowns. The conditions amount to the linear (!) system

$$\begin{aligned} p(1) &= a_0 + a_1 + a_2 + a_3 && = 0 \\ p(2) &= a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 + a_3 \cdot 2^3 && = 3. \end{aligned}$$

This can be rewritten as

$$\begin{aligned} a_0 + a_1 + a_2 + a_3 &= 0 \\ a_0 + 2a_1 + 4a_2 + 8a_3 &= 3. \end{aligned}$$

Using Gaussian elimination to solve this: the associated matrix is

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 4 & 8 & 3 \end{array}\right).$$

Subtracting the first from the second row gives

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 7 & 3 \end{array}\right).$$

Subtracting the second from the first yields a reduced row echelon matrix:

$$\left(\begin{array}{cccc|c} 1 & 0 & -2 & -6 & -3 \\ 0 & 1 & 3 & 7 & 3 \end{array}\right).$$

The variables a_0 and a_1 correspond to the leading 1's, the variables a_2 and a_3 are therefore free variables. Thus, there are infintely many solutions. One solution, for $a_2 = a_3 = 0$ is

$$a_0 = -3, \ a_1 = 3,$$

so that

$$p(x) = -3 + 3x$$

is a solution to the problem. Another solution would be $a_2 = a_3 = 1$, which gives $a_1 = -7$ and $a_0 = 5$, i.e.,

$$p(x) = 5 - 7x + x^2 + x^3.$$

C.3 Vector spaces

Solution of Exercise 3.13: A linear combination of A and B is of the form

$$\alpha A + \beta B = \alpha \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} + \beta \begin{pmatrix} 3 & 2 \\ 3 & 5 \end{pmatrix}$$

with $\alpha, \beta \in \mathbf{R}$. Computing the left hand side, we need to find α and β such that

$$\begin{pmatrix} \alpha & \alpha \\ 2\alpha & 2\alpha \end{pmatrix} + \begin{pmatrix} 3\beta & 2\beta \\ 3\beta & 5\beta \end{pmatrix} = \begin{pmatrix} \alpha + 3\beta & \alpha + 2\beta \\ 2\alpha + 3\beta & 2\alpha + 5\beta \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 2 & 4 \end{pmatrix}.$$

Comparing the entries of the matrix, this gives the linear system

$$\begin{aligned} \alpha + 3\beta &= -1 \\ \alpha + 2\beta &= 0 \\ 2\alpha + 3\beta &= 2 \\ 2\alpha + 5\beta &= 4. \end{aligned}$$

The second gives $\alpha = -2\beta$, inserting into the first gives $-2\beta + 3\beta = -1$, which means $\beta = -1$. However, inserting into the third equation gives $-4\beta + 3\beta = 2$, so that $\beta = -2$, contradicting the previous equation. Thus, there is no solution, so C is *not* a linear combination of A and B.

Solution of Exercise 3.15: The system $x + y + z + t = 0$ corresponds to the matrix

$$(1\ 1\ 1\ 1).$$

This matrix is already in reduced row echelon form: the leading one is for the variable x, the variables y, z, t are free variables. Thus,

$$S = \{(-\alpha - \beta - \gamma, \alpha, \beta, \gamma) \mid \alpha, \beta, \gamma \in \mathbf{R}\}.$$

We have

$$\begin{aligned} S &= \{(-\alpha, \alpha, 0, 0) + (-\beta, 0, \beta, 0) + (-\gamma, 0, 0, \gamma) \mid \alpha, \beta, \gamma \in \mathbf{R}\} \\ &= \{\alpha(-1, 1, 0, 0) + \beta(-1, 0, 1, 0) + \gamma(-1, 0, 0, 1) \mid \alpha, \beta, \gamma \in \mathbf{R}\} \\ &= L((-1, 1, 0, 0), (-1, 0, 1, 0), (-1, 0, 0, 1)). \end{aligned}$$

Solution of Exercise 3.16: By definition, S consists of all the linear combinations of the three given vectors. These can be written as

$$a(1, -1, 0, 1) + b(2, 1, -2, 0) + c(0, 0, 1, 1) = (a+2b, -a+b, -2b+c, a+c)$$

for arbitrary $a, b, c \in \mathbf{R}$. The intersection is given by vectors as above satisfying the linear system determining T, i.e.,

$$\begin{aligned} x_1 &= a + 2b \\ x_2 &= -a + b \\ x_3 &= -2b + c \\ x_4 &= a + c \end{aligned}$$

such that

$$\begin{aligned} 2(a + 2b) - (-a + b) - 3(a + c) &= 0 \\ 2(a + 2b) + (-2b + c) + (a + c) &= 0. \end{aligned}$$

Simplifying these equations gives

$$\begin{aligned} 3b - 3c &= 0 \\ 3a + 2b + 2c &= 0. \end{aligned}$$

Thus $b = c$ and $3a + 4c = 0$, i.e., $a = -\frac{4}{3}c$, and c is a free variable. (Alternatively, the above system is associated to the matrix $\begin{pmatrix} 0 & 3 & -3 \\ 3 & 2 & 2 \end{pmatrix}$, which can be brought into reduced row echelon form.) Thus,

$$\begin{aligned} S \cap T &= \{-\frac{4}{3}c(1,-1,0,1) + c(2,1,-2,0) + c(0,0,1,1) \mid c \in \mathbf{R}\} \\ &= \{c\left((-\frac{4}{3},\frac{4}{3},0,-\frac{4}{3}) + (2,1,-2,0) + (0,0,1,1)\right) \mid c \in \mathbf{R}\} \\ &= \{c(\frac{2}{3},\frac{7}{3},-1,-\frac{1}{3}) \mid c \in \mathbf{R}\} \\ &= L((\frac{2}{3},\frac{7}{3},-1,-\frac{1}{3})). \end{aligned}$$

Solution of Exercise 3.25: We have to find a vector $v \in W_1$ that is also contained in W_2. This means that

$$v = a(1,0,1) + b(2,1,0) = (a + 2b, b, a) \tag{C.1}$$

for some $a, b \in \mathbf{R}$ and at the same time

$$v = \alpha(-1,-1,1) + \beta(0,3,0) = (-\alpha, -\alpha + 3\beta, \alpha)$$

for some $\alpha, \beta \in \mathbf{R}$. Comparing the two vectors gives the following linear system, where a, b, α, β are the unknowns:

$$\begin{aligned} a + 2b &= -\alpha \\ b &= -\alpha + 3\beta \\ a &= \alpha. \end{aligned}$$

We solve this system: the last equation gives $a = \alpha$ and, from the first equation, $b = -\alpha$. The second equation implies $\beta = 0$. There is no condition on α, this $\alpha = r$ for an arbitrary real number $r \in \mathbf{R}$.

Instead of solving the above system by hand, we may also use Gaussian elimination to solve this linear system. The matrix is the following (where the columns are for a, b, α, β, in that order):

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -3 \\ 1 & 0 & -1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -3 \\ 0 & -2 & -2 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & -6 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The three leading ones are for the variables a, b, β, and α is a free variable, so let $\alpha = r$, where $r \in \mathbf{R}$ is an arbitrary real number. This gives again $\beta = 0$, $b + r - 3\beta = 0$, so that $b = -r$ and $a = r$.

Thus the intersection $W_1 \cap W_2$ consists of the vectors

$$v = \alpha(1,0,1)+(-\alpha)(2,1,0) = \alpha(-1,-1,1)+0(0,3,0) = (-\alpha,-\alpha,\alpha).$$

Thus,

$$W_1 \cap W_2 = L((1,-1,1)),$$

so a basis of $W_1 \cap W_2$ consists of (the single vector) $(1,-1,1)$, and in particular

$$\dim W_1 \cap W_2 = 1.$$

We now consider $W_1 + W_2$. According to Definition 3.34,

$$W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\},$$

i.e., of arbitrary sums whose two summands are in W_1, respectively W_2.

As was noted in the proof of Corollary 3.72, if $V_1 = L(v_1, \dots, v_n)$ and $V_2 = L(w_1, \dots, w_m)$ are two subspaces of a vector space V, then the sum

$$V_1 + V_2 = L(v_1, \dots, v_n, w_1, \dots, w_m).$$

For the subspaces W_1, W_2 above, this means that we determine the span

$$L(\underbrace{(1,0,1)}_{v_1}, \underbrace{(2,1,0)}_{v_2}, \underbrace{(-1,-1,1)}_{w_1}, \underbrace{(0,3,0)}_{w_2}).$$

By Definition 3.58(1), we obtain a basis of $W_1 + W_2$ by (possibly) removing several of these four vectors. To determine which ones these are, we apply Method 3.53 and Method 3.44. The matrix built out of the four vectors is

$$\begin{pmatrix} v_1 \\ v_2 \\ w_1 \\ w_2 \end{pmatrix} = \left(\begin{array}{ccc} 1 & 0 & 1 \\ 2 & 1 & 0 \\ \hline -1 & -1 & 1 \\ 0 & 3 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & -2 \\ \hline 0 & -1 & 2 \\ 0 & 3 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & -2 \\ \hline \underline{0} & \underline{0} & \underline{0} \\ 0 & 0 & 6 \end{array}\right).$$

Note that in this process we only added multiples of some rows to another row, but did not interchange any rows. Since we have the zero vector (underlined) in the third row, the vector w_1 is a linear combination of v_1 and v_2. The vectors v_1, v_2, w_2 are however linearly independent. Thus, they form a basis of $W_1 + W_2$. In particular, $\dim(W_1 + W_2) = 3$.

An alternative way to determine at least the dimension of W_1+W_2 is to use Theorem 3.73:

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Using again Method 3.53, one can check that v_1, v_2 is a basis of W_1, so that $\dim W_1 = 2$ and similarly that w_1, w_2 form a basis of W_2, so that $\dim W_2 = 2$. Thus, using the first part of the exercise, we confirm $\dim(W_1 + W_2) = 3$.

Solution of Exercise 3.27: We will show that v_1, v_2, v_3 are linearly independent (in $\mathbf{R}^4$ and therefore also in the subspace W) and therefore form a basis of W. We use Method 3.53:

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This matrix has three leading ones, so the vectors are linearly independent as claimed.

We "guess" $v = (1, 2, 3, 4)$ and check that these vectors v_1, v_2, v_3, v are linearly independent. By Lemma 3.51, this will then imply that v is not a linear combination of the other vectors, so that $W \subsetneq$

$L(v_1, v_2, v_3, v)$. We use Method 3.53:

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 2 & 2 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 1 & 1 & 2 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} \underline{1} & 0 & 1 & 0 \\ 0 & 0 & 0 & \underline{4} \\ 0 & 0 & \underline{1} & 3 \\ 0 & \underline{1} & 1 & 2 \end{pmatrix}$$

After dividing the second row by 4, we can interchange rows and get a row echelon matrix with four leading ones (underlined). Thus, v_1, v_2, v_3, v are linearly independent. Therefore, they form in fact a basis of $\mathbf{R}^4$, and we know by Definition 3.58(3) that therefore

$$W \subsetneq \mathbf{R}^4 = L(v_1, v_2, v_3, v).$$

Remark C.2. A more systematic way of solving the second part of the exercise, without guessing, is to use Definition 3.58: we can take the standard basis of $\mathbf{R}^4$, and for (at least) one of the four standard basis vectors e_1, e_2, e_3, e_4 we will have that this standard basis vector together with v_1, v_2, v_3 form a basis of $\mathbf{R}^4$. We can then use Method 3.53 to see that, for example, v_1, v_2, v_3, e_1 are linearly independent and therefore form a basis of $\mathbf{R}^4$, so that in particular $W \subsetneq L(v_1, v_2, v_3, e_1)$.

Solution of Exercise 3.28: We bring the matrix formed by these vectors in row-echelon form:

$$\begin{pmatrix} 1 & 0 & -1 & 2 \\ 1 & 0 & 0 & 1 \\ 2 & 0 & -1 & 3 \\ 4 & t & -2 & 6 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & t & 2 & -2 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & t & 2 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & t & 2 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

If $t \neq 0$, we can divide by t, which gives a matrix with three leading ones. Thus, the space U_t which is spanned by these vectors has dimension 3 in this case. If $t = 0$, we continue simplifying the matrix into row echelon form:

$$\begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & t & 2 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This has two leading ones, thus $\dim U_t = 2$ in this case.

We now consider $t = 1$. The subspace $U := U_1$ then has a basis consisting of the non-zero rows if the matrix above, i.e., it has a basis consisting of the vectors

$$(1, 0, -1, 2), (0, 1, 2, -2), (0, 0, 1, -1).$$

In order to determine a basis of W, we form the matrix associated to these homogeneous equations, which is

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & -1 & -1 & -3 \end{pmatrix}.$$

This has two columns not having a leading one, namely the last two. These are the free variables, say $x_3 = a$, $x_4 = b$ for $a, b \in \mathbf{R}$. To determine a basis of W, we therefore have to consider the system

$$\begin{aligned} x_1 + x_2 + a &= 0 \\ x_2 + a + 3b &= 0 \end{aligned}$$

This gives $x_2 = -a - 3b$, and $x_1 - 3b = 0$ so that $x_1 = 3b$. Thus, a basis of W is given by the two vectors

$$(0, -1, 1, 0) \text{ and } (3, -3, 0, 1).$$

In order to determine $U \cap W$, consider a generic vector of U, i.e., one of the form

$$\begin{aligned} v &= a(1, 0, -1, 2) + b(0, 1, 2, -2) + c(0, 0, 1, -1) \\ &= (a, b, -a + 2b + c, 2a - 2b - c). \end{aligned}$$

We require it to satisfy the equations describing W:

$$\begin{aligned} a + b + (-a + 2b + c) &= 0 \\ a - 3(2a - 2b - c) &= 0. \end{aligned}$$

Simplifying these expressions gives the system

$$\begin{aligned} 3b + c &= 0 \\ -5a + 6b + 3c &= 0. \end{aligned}$$

Therefore $c = -3b$, plugging this into the second equation gives, after simplifying, $-5a - 3b = 0$ or $a = -\frac{3}{5}b$. Thus, our vector $v \in U$ belongs to W precisely if it can be written as

$$\begin{aligned} -\frac{3}{5}b(1, 0, -1, 2) + b(0, 1, 2, -2) + (-3b)(0, 0, 1, -1) &= (-\frac{3}{b}b, b, \frac{3}{5}b + 2b - 3b, -\frac{6}{5}b - 2b + 3 \\ &= b(-\frac{3}{5}, 1, -\frac{2}{5}, -\frac{1}{5}), \end{aligned}$$

where $b \in \mathbf{R}$ is arbitrary. Thus, a basis of $U \cap W$ is this vector

$$(-\frac{3}{5}, 1, -\frac{2}{5}, -\frac{1}{5}).$$

In particular, $\dim U \cap W = 1$.

Solution of Exercise 3.30: The map f is given by multiplication with the matrix, which we bring to row echelon form by elementary row operations

$$\begin{pmatrix} 0 & -1 & 2 & 2 \\ 1 & 0 & -1 & 2 \\ -5 & 1 & 2 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 0 & -1 & 2 & 0 \\ -3 & 6 & 0 & 4 \\ -8 & 1 & 2 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -2 & 1 & 2 & -3 \end{pmatrix}$$

It has 2 leading ones in the first two columns, so $\operatorname{im} f = L((0,2,-1,1),(-1,1,2,2))$. The third variable z is a free variable, and $\ker f = L(-1,2,1)$.

The space W defined by the equation $x_2 - 3x_3 = 0$ is the kernel of the matrix $\begin{pmatrix} 0 \\ 1 \\ -3 \\ 0 \end{pmatrix}$, which gives (for example) three free variables x_1, x_3, x_4, and accordingly a basis consisting of $(1,0,0,0),(0,3,1,0),(0,0,0,1)$. We have $\dim W = 3$.

An element in $U = \operatorname{im} f$ is a linear combination of the first two column vectors, so of the form $(-b, 2a+b, -a+2b, a+2b)$, for arbitrary $a, b \in \mathbf{R}$. This lies in W precisely if $2a+b-3(-a+2b) = 0$, i.e., if $5a - 5b = 0$ or if $a = b$. A basis of $U \cap W$ is then given by choosing, say, $a = 1$, and $U \cap W = L(-1,3,1,3)$.

We know $\dim U+W = \dim U + \dim W - \dim(U \cap W) = 2+3-1 = 4$. So, for example, the three basis vectors of W, together with any $u \in U \backslash W$ will form a basis. Thus, $(1,0,0,0),(0,3,1,0),(0,0,0,1),(0,2,-1,1)$ forms a basis of $U + W$. (Note the last vector is not in W since its does not satisfy the equation $x_2 - 3x_3 = 0$.)

In order to check when $(a,4,3,b)$ is in the image, we form the matrix corresponding to the linear system:

$$\left(\begin{array}{cc|c} 0 & -1 & a \\ 2 & 1 & 4 \\ -1 & 2 & 3 \\ 1 & 2 & b \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 0 & -1 & a \\ 0 & -3 & 4-2b \\ 0 & 4 & b+3 \\ 1 & 2 & b \end{array}\right).$$

This shows that $4 - 2b = 3a$ and $4(4-2b) + 3(b+3) = 0$, i.e., $25 - 5b = 0$, so $b = 5$ and $a = -2$.

C.4 Linear maps

Solution of Exercise 4.9: Recall that $\ker f = \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbf{R}^2 \mid f(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\}$. Thus a vector is in the kernel precisely if it is a solution of the homongeneous system

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ x_2 &= 0 \\ 3x_1 + 5x_2 &= 0. \end{aligned}$$

Solving this system gives $x_2 = 0$, then $x_1 = 0$. Thus, $\ker f = \{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\}$.

For the second question, recall that $\operatorname{im} f$ consists precisely of those vectors $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ that are of the form $\begin{pmatrix} a \\ b \\ c \end{pmatrix} = f(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix})$ for some $x_1, x_2 \in \mathbf{R}$. Thus, the question amounts to this: do there exist $x_1, x_2 \in \mathbf{R}$ with

$$f(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} x_1 + 2x_2 \\ x_2 \\ 3x_1 + 5x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}?$$

Again, this leads to a linear system:

$$\begin{aligned} x_1 + 2x_2 &= 1 \\ x_2 &= 0 \\ 3x_1 + 5x_2 &= 3. \end{aligned}$$

The first two equations give $x_2 = 0$, $x_1 = 1$. This also satisfies the last equation, so the vector $\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$ is indeed in the image, because $\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} = f(\begin{pmatrix} 1 \\ 0 \end{pmatrix})$.

Solution of Exercise 4.10: To determine the kernel of f, one has to solve the homogeneous system

$$\begin{aligned} 2x_2 - x_2 + x_3 + x_4 &= 0 \\ 5x_2 - 3x_3 - 5x_4 &= 0 \\ 3x_1 - 4x_2 + 3x_3 + 4x_4 &= 0. \end{aligned}$$

For the second task, one has to solve the non-homogeneous system

$$\begin{aligned} 2x_2 - x_2 + x_3 + x_4 &= 1 \\ 5x_2 - 3x_3 - 5x_4 &= -3 \\ 3x_1 - 4x_2 + 3x_3 + 4x_4 &= 3. \end{aligned}$$

This solution set is *not* a subspace, since the zero vector $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ is not a solution for the system: the left hand side of all three equations is 0, while the right ones are not.

Solution of Exercise 4.11: We compute the rank using Gaussian elimination:

$$\begin{aligned} A_t = \begin{pmatrix} 1 & 3 & -1 & 2 \\ 1 & 5 & 1 & 1 \\ 2 & 4 & t & 5 \end{pmatrix} &\rightsquigarrow \begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 2 & 2 & -1 \\ 0 & -2 & t+2 & 1 \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 2 & 2 & -1 \\ 0 & 0 & t+4 & 0 \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 1 & 1 & -\frac{1}{2} \\ 0 & 0 & t+4 & 0 \end{pmatrix} \end{aligned}$$

If $t \neq -4$, then we can further divide the last row by $t+4 (\neq 0)$, and the rank is then 3. For $t = -4$, the rank is 2.

For the second task, the system we are considering here is

$$\begin{aligned} x_1 + 3x_2 - x_3 + 2x_4 &= 1 \\ x_1 + 5x_2 + x_3 + x_4 &= \alpha \\ 2x_1 + 4x_2 - 4x_3 + 5x_4 &= 0. \end{aligned}$$

For the last task: the rank of A_t is at most 3. Thus, $\dim \operatorname{im} f \leqslant 3$, and therefore

$$\dim \ker = \dim \mathbf{R}^4 - \dim \operatorname{im} f \geqslant 4 - 3 = 1.$$

This means that, for all t, the kernel of f is not just consisting of the zero vector, hence the answer to the question is no.

Solution of Exercise 4.14: To determine a basis of $\ker f$ and of $\operatorname{im} f$, we bring A into row echelon form:

$$A = \begin{pmatrix} 2 & -1 & -\frac{5}{2} & 1 \\ -1 & 0 & 1 & -\frac{1}{2} \\ 1 & 1 & \frac{-}{1}2 & \frac{1}{2} \\ 0 & 2 & 1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 0 & -3 & -\frac{3}{2} & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 0 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & -3 & -\frac{3}{2} & 0 \\ 0 & 2 & 1 & 0 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} =: B.$$

According to Proposition 4.32, $\operatorname{rk} A = \dim \operatorname{im} f$ equals the number of columns (of B) with a leading 1, i.e., $\dim \operatorname{im} f = 2$. A basis of $\operatorname{im} f$ is given by the columns of A corresponding to these columns of B, i.e., the first two columns. Thus, a basis of $\operatorname{im} f$ is given by the two vectors $(2, -1, 1, 0)$ and $(-1, 0, 1, 2)$. To determine a basis of $\ker f$, the third and fourth columns correspond to a free variable, i.e., we can choose $x_3 = a$, $x_4 = b$ with $a, b \in \mathbf{R}$ arbitrary. We obtain the equations

$$\begin{aligned} x_1 + x_2 - \frac{1}{2}a + \frac{1}{2}b &= 0 \\ x_2 + \frac{1}{2}a &= 0. \end{aligned}$$

This gives $x_2 = -\frac{1}{2}a$ and $x_1 = a - \frac{1}{2}b$. Therefore,

$$\begin{aligned} \ker f &= \{(a - \frac{1}{2}b, -\frac{1}{2}a, a, b \mid a, b \in \mathbf{R}\} \\ &= \{a(1, -\frac{1}{2}, 1, 0) + b(-\frac{1}{2}, 0, 0, 1) \mid a, b \in \mathbf{R}\} \\ &= L((1, -\frac{1}{2}, 1, 0), (-\frac{1}{2}, 0, 0, 1)). \end{aligned}$$

These two vectors form a basis of $\ker f$.

In order to determine im $f \cap \ker f$, we need to consider elements of

$$\text{im } f = \{a(2, -1, 1, 0) + b(-1, 0, 1, 2) \mid a, b \in \mathbf{R}\}$$

that also belong to the kernel, i.e., the vector $(2a - b, -a, a + b, 2b)$ must lie in $\ker f$. This means that

$$A \begin{pmatrix} 2a - b \\ -a \\ a + b \\ 2b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Again we use the row-echelon form of A, computed above. So this system is equivalent to

$$B \begin{pmatrix} 2a - b \\ -a \\ a + b \\ 2b \end{pmatrix} = \begin{pmatrix} 1 & 1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2a - b \\ -a \\ a + b \\ 2b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This gives the two equations

$$\begin{aligned} (2a - b) + (-a) - \frac{1}{2}(a + b) + \frac{1}{2}(2b) &= 0 \\ -a + \frac{1}{2}(a + b) &= 0. \end{aligned}$$

This simplifies to

$$\begin{aligned} \frac{a}{2} - \frac{b}{2} &= 0 \\ -\frac{a}{2} + \frac{b}{2} &= 0. \end{aligned}$$

This is equivalent to the condition $a = b$. Therefore,

$$\begin{aligned} \ker f \cap \operatorname{im} f &= \{a(2, -1, 1, 0) + a(-1, 0, 1, 2) \mid a \in \mathbf{R}\} \\ &= \{a(1, -1, 2, 2) \mid a \in \mathbf{R}\} \\ &= L((1, -1, 2, 2)). \end{aligned}$$

That is, the vector $(1, -1, 2, 2)$ is a basis of $\ker f \cap \operatorname{im} f$.

Solution of Exercise 4.16: We have $v_1 = (1, 1, 0)$, and compute

$$v_2 = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix},$$

$$v_3 = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 0 \end{pmatrix}.$$

In order to confirm that they form a basis, we apply Method 3.44 and Method 3.53 by forming the associated matrix and bringing it into row echelon form:

$$\begin{aligned} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 5 & 0 \end{pmatrix} &\rightsquigarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ -4 & 0 & 0 \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 4 & 0 \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & -8 \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

This matrix has three leading ones, so that the vectors do form a basis.

We compute $v_4 = f(v_3) = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 10 \end{pmatrix}$.

The equation $v_4 = a_1 v_1 + a_2 v_2 + a_3 v_3$ is the linear system

$$\begin{aligned} a_1 + a_2 + a_3 &= -3 \\ a_1 + a_2 + 5a_3 &= 1 \\ 2a_2 &= 10. \end{aligned}$$

We solve this: the last equation gives $a_2 = 5$, which leads to

$$\begin{aligned} a_1 + 5 + a_3 &= -3 \\ a_1 + 5 + 5a_3 &= 1. \end{aligned}$$

Therefore

$$\begin{aligned} a_1 + a_3 &= -8 \\ a_1 + 5a_3 &= -4. \end{aligned}$$

This can be solved to $a_3 = 1$ and $a_1 = -9$. Thus, $(a_1, a_2, a_3) = (1, 5, -9)$ are the coordinates of v_4 in the basis v_1, v_2, v_3.

We now determine the matrix of f with respect to the basis v_1, v_2, v_3 (both in the domain and the codomain of f). We therefore write each $f(v_i)$ as a linear combination of these three vectors:

$$\begin{aligned} f(v_1) &= v_2 = 0v_1 + 1v_2 + 0v_3 \\ f(v_2) &= v_3 = 0v_1 + 0v_2 + 1v_3 \\ f(v_3) &= v_4 = -9v_1 + 5v_2 + v_3. \end{aligned}$$

According to Proposition 4.43, the matrix of f with respect to v_1, v_2, v_3 is

$$\begin{pmatrix} 0 & 0 & -9 \\ 1 & 0 & 5 \\ 0 & 1 & 1 \end{pmatrix}.$$

Solution of Exercise 4.17: The matrix for f is

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 \end{pmatrix}$$

To compute the image of f, we bring this matrix into row-echelon form:

$$A \rightsquigarrow \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

After multiplying the first two rows with -1, we get a matrix with three leading ones. Therefore $\dim \operatorname{im} f = 3$, which implies that $\operatorname{im} f = \mathbf{R}^3$. This tells us that $\dim \ker f = 1$, so $\ker f$ is generated by any non-zero vector in it. An example of such a vector is $\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$, which therefore constitutes a basis of $\ker f$.

Solution of Exercise 4.18:

(1) f_2 and f_3 are not linear ($f(a+b) \neq f(a) + f(b)$ for most a and b). The remaining ones are linear.

(2)

$$\begin{aligned} \operatorname{im}\,(f_1) &= L(\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}) \\ \operatorname{im}\,(f_2) &= \mathbf{R}^2 \\ \operatorname{im}\,(f_3) &= \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbf{R}^2 | x_1 \geqslant 0\} \\ \operatorname{im}\,(f_4) &= \mathbf{R}^3 \\ \operatorname{im}\,(f_5) &= \mathbf{R}^3 \end{aligned}$$

For the linear maps f_1, f_4, f_5, this can be checked by forming the matrix of these maps with respect to the standard bases and bringing it into row echelon form.

(3) According to the rank-nullity theorem, we see that f_1 is injective, f_4 is bijective (i.e. sujective e injective) and f_5 is bijective.

Solution of Exercise 4.27: We write any vector $v = (a, b) = a(1, 0) + b(0, 1) = ae_1 + be_2$ in terms of v_1, v_2:

$$v = \alpha v_1 + \beta v_2 = \alpha(1, -3) + \beta(2, 1).$$

As an example, if $v = (0, 6)$, then $(0, 6) = \alpha(1, -3) + \beta(2, 1)$ gives the linear system

$$\begin{aligned} 0 &= \alpha + 2\beta \\ 6 &= -3\alpha + \beta. \end{aligned}$$

This can be solved as $\beta = \frac{6}{7}$ and $\alpha = -\frac{12}{7}$. Thus, we can write

$$v = (0, 6)_{(e_1,e_2)} = (-\frac{12}{7}, \frac{6}{7})_{(v_1,v_2)},$$

where the subscripts indicate that the coordinates are with respect to the standard basis, resp. to the basis v_1, v_2.

We now determine the base change matrix. We have to write the matrix of the identity map in terms of the standard basis in the domain, and the basis v_1, v_2 in the codomain:

$$\begin{aligned} e_1 = (1, 0) &\mapsto \mathrm{id}(e_1) = (1, 0) = \alpha_1 v_1 + \alpha_2 v_2 \\ e_2 = (0, 1) &\mapsto \mathrm{id}(e_2) = (0, 1) = \beta_1 v_1 + \beta_2 v_2. \end{aligned}$$

The base change matrix is then the matrix

$$\begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix}.$$

We can compute α_1 etc. similarly as above:

$$\begin{aligned} 1 &= \alpha_1 + 2\alpha_2 \\ 0 &= -3\alpha_1 + \alpha_2 \end{aligned}$$

We solve this as $\alpha_1 = \frac{1}{7}$, $\alpha_2 = \frac{3}{7}$. As for β_1, β_2, the relevant system is

$$\begin{aligned} 0 &= \beta_1 + 2\beta_2 \\ 1 &= -3\beta_1 + \beta_2 \end{aligned}$$

whose solution is $\beta_1 = -\frac{2}{7}$, $\beta_2 = \frac{1}{7}$. Therefore, the base change matrix is

$$H = \begin{pmatrix} \frac{1}{7} & -\frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} \end{pmatrix}.$$

We compute the coordinates of $(2,-5)$ in terms of the basis v_1, v_2:

$$H\begin{pmatrix} 2 \\ -5 \end{pmatrix} = \begin{pmatrix} \frac{1}{7} & -\frac{2}{7} \\ \frac{3}{7} & \frac{1}{7} \end{pmatrix}\begin{pmatrix} 2 \\ -5 \end{pmatrix} = \begin{pmatrix} \frac{12}{7} \\ \frac{1}{7} \end{pmatrix}.$$

Thus, the coordinates of $(2,-5)$ with respect to the basis v_1, v_2 is $(\frac{12}{7}, \frac{1}{7})$.

Solution of Exercise 4.28: We have, for example,

$$\mathrm{id}(v_1) = v_1 = (1,0,-1) = 1\cdot e_1 + 0\cdot e_2 + (-1)\cdot e_3.$$

Likewise for v_2 and v_3. Therefore, the base change matrix is

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 7 \end{pmatrix}.$$

Solution of Exercise 4.30: We follow the given hint, and first compute H. We have

$$\begin{aligned} v_1 &\mapsto v_1 = (1,-1) = e_1 - e_2, \\ v_2 &\mapsto v_2 = (3,-1) = 3e_1 - e_2. \end{aligned}$$

Thus $H = \begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix}$.

We now compute the base change matrix K from the standard basis to the basis $\underline{v} = \{v_1, v_2, v_3\}$:

$$e_1 = (1,0,0) \mapsto (1,0,0) = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3.$$

Plugging in the values of v_1, v_2, v_3, this gives the linear system

$$\begin{aligned} 1 &= \alpha_1 + 2\alpha_2 - \alpha_3 \\ 0 &= \alpha_2 - \alpha_3 \\ 0 &= \alpha_1 + \alpha_2 - \alpha_3. \end{aligned}$$

This has the solution $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = 1$. Similarly, if instead of $e_1 = (1,0,0)$, we consider $e_2 = (0,1,0)$, the constants in the above system change accordingly to 0, resp. 1, resp. 0 in the three equations above. The solution is then $\alpha_1 = 1$, $\alpha_2 = -2$, $\alpha_3 = -3$.

Similarly, for e_3, we obtain the solution $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = -1$. Hence

$$K = \begin{pmatrix} 0 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -3 & -1 \end{pmatrix}.$$

We compute

$$\begin{aligned} KAH &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -3 & -1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ -3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -3 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 5 \\ -1 & -1 \\ -4 & -10 \end{pmatrix} \\ &= \begin{pmatrix} -5 & -11 \\ 7 & 17 \\ 8 & 18 \end{pmatrix}. \end{aligned}$$

Solution of Exercise 4.32: We follow the solution of Exercise 4.30, see above:

$$\mathbf{R}^2_{\underline{v}} \xrightarrow[H]{\text{id}} \mathbf{R}^2_{\underline{e}} \xrightarrow[A]{f} \mathbf{R}^2_{\underline{e}} \xrightarrow[K]{\text{id}} \mathbf{R}^2_{\underline{v}}.$$

We have $H = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. One computes $K = \begin{pmatrix} 2 \\ -1 \end{pmatrix} -11$ and then

$$\begin{aligned} KAH &= \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 6 & -1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 4 \\ 5 & 8 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}. \end{aligned}$$

Solution of Exercise 4.33: For the given matrix A, the system $Ax = x$ is written out like so:

$$\begin{aligned} x_3 &= x_1 \\ x_2 &= x_2 \\ x_1 &= x_3. \end{aligned}$$

The second equation holds for all x, and the first is equivalent to the third. Therefore, the system is equivalent to the one consisting of the single equation

$$x_1 - x_3 = 0.$$

This corresponds to the system

$$Bx = x,$$

where $B = (1\ 0\ -1)$ is the corresponding 1×3-matrix. This matrix has rank 1, so that the solution space is two-dimensional, given by

$$L((1,0,1),(0,1,0)).$$

Solution of Exercise 4.34: We write out the given system

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 5 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

as

$$\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} - \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \right] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This simplifies to

$$\underbrace{\begin{pmatrix} -4 & 0 & 0 \\ 0 & -1 & 2 \\ 0 & 2 & -4 \end{pmatrix}}_{=:B} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which can be solved by bringing B into row-echelon form:

$$B \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus $x_1 = 0$, $x_2 = 2x_3$ and x_3 is a free variable. Thus the solution space of the given system is

$$L((0,2,1)).$$

Solution of Exercise 4.35: We proceed the same way as for Exercise 4.30 (see its solution above):

$$\mathbf{R}^2_{\underline{v}} \xrightarrow[H]{\mathrm{id}} \mathbf{R}^2_{\underline{e}} \overset{A}{f}\!\longrightarrow \mathbf{R}^2_{\underline{e}} \xrightarrow[H^{-1}]{\mathrm{id}} \mathbf{R}^2_{\underline{v}}.$$

Here H is the base change matrix from $\underline{v}$ to the standard basis $\underline{e} = \{e_1 = (1,0), e_2 = (0,1)\}$. The base change matrix from $\underline{e}$ to $\underline{v}$ is then H^{-1}. From the given vectors we have $H = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$. We compute the inverse H^{-1} using Theorem 4.80:

$$\left(\begin{array}{cc|cc} 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & -2 & 1 & -2 \end{array}\right)$$
$$\rightsquigarrow \left(\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array}\right)$$
$$= (\mathrm{id} \mid H^{-1}).$$

Thus, $H^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix}$. Therefore the requested matrix of f is

$$H^{-1}AH = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 15 & - \end{pmatrix} 1$$
$$= \begin{pmatrix} 3 & 0 \\ 12 & -1 \end{pmatrix}.$$

$\blacksquare$

Solution of Exercise 4.37: This can be done as for Exercise 4.35 above. The final solution is

$$\begin{pmatrix} -2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Solution of Exercise 4.38: One computes this solution space as

$$L((-1,2,3)) = L((1,-2,-3)).$$

According to Theorem 3.67(2), this vector $l_1 = (1,-2,-3)$ can be completed to a basis of $\mathbf{R}^3$ by picking *any* basis v_1, v_2, v_3 of $\mathbf{R}^3$. Then it is possible to find two of these three vectors which together with l_1 will form a basis of $\mathbf{R}^3$. We pick the standard basis, $v_1 = e_2$, $v_2 = e_2$ and $v_3 = e_3$. We check that l_1, e_2, e_3 form a basis. Indeed, the matrix whose rows are these vectors,

$$\begin{pmatrix} 1 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

is a row-echelon matrix with three leading ones, so its rank is 3.

We compute the matrix of f with respect to this basis $\underline{v} = \{l_1, e_2, e_3\}$:

$$\mathbf{R}^3_{\underline{v}} \xrightarrow[H]{\text{id}} \mathbf{R}^3_{\underline{e}} \xrightarrow[A]{f} \mathbf{R}^3_{\underline{e}} \xrightarrow[H^{-1}]{\text{id}} \mathbf{R}^3_{\underline{v}}.$$

We have

$$H = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}.$$

We compute the inverse using Theorem 4.80:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 1 & 0 \\ -3 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 0 & 1 \end{array}\right) = (\text{id} \mid H^{-1}).$$

Hence the matrix for f with respect to the basis $\underline{v}$ is

$$H^{-1}AH = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & 3 \end{pmatrix}.$$

Remark C.3. The choice of the two vectors e_2 and e_3, in addition to l_1 above, is arbitrary. To begin with, one may choose a different basis (other than the standard basis) to complete l_1 to a basis. Even if one takes the standard basis, for this particular value of l_1, *any* two of the three vectors e_1, e_2, e_3 together with l_1 would form a basis. The resulting base change matrix H will then be different, and also the result $H^{-1}AH$ will be different.

Solution of Exercise 4.39: The vectors

$$\underline{v} = \{v_1 = (1, 0, 1), v_2 = (0, 3, -1), v_3 = (0, 0, 1)\}$$

form a basis of $\mathbf{R}^3$ since the corresponding matrix $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{pmatrix}$ has rank 3. Hence we can compute the matrix of f with respect to that basis as follows:

$$\begin{aligned} f(v_1) &= 0 & &= 0v_1 + 0v_2 + 0v_3 \\ f(v_2) &= v_2 & &= 0v_1 + 1v_2 + 0v_3 \\ f(v_3) &= (0,0,2) = 2v_3 & &= 0v_1 + 0v_2 + 2v_3. \end{aligned}$$

Therefore the matrix of f with respect to that basis is

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

As before, we compute that matrix of f with respect to the standard basis using the method above:

$$\mathbf{R}^3_{\underline{e}} \xrightarrow[K]{\text{id}} \mathbf{R}^3_{\underline{v}} \xrightarrow[A]{f} \mathbf{R}^3_{\underline{v}} \xrightarrow[K^{-1}]{\text{id}} \mathbf{R}^3_{\underline{e}}.$$

The base change matrix K is easily read off:

$$K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

We compute the inverse K^{-1}:

$$\begin{aligned} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 \end{array}\right) &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right) \\ &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right) \\ &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -1 & \frac{1}{3} & 1 \end{array}\right) \\ &= (\text{id} \mid K^{-1}). \end{aligned}$$

Therefore the requested matrix is the product

$$\begin{aligned} KAK^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ -1 & \frac{1}{3} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ -2 & \frac{2}{3} & 2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & \frac{1}{3} & 2 \end{pmatrix}. \end{aligned}$$

Solution of Exercise 4.41: It is convenient to observe

$$5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} 5 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Thus the given system can be rewritten as

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

By Lemma 4.59, this is the same as the system

$$\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix} - \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \right] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The left hand matrix equals $\begin{pmatrix} -4 & 0 & 0 \\ 0 & -1 & 2 \\ 0 & 2 & -4 \end{pmatrix}$. From here, one can use the standard method to solve the linear system. (As a forecast to §5, one can note that the determinant of the latter matrix is 0, so that the matrix is not invertible. Hence the system above has non-zero solutions.)

Solution of Exercise 4.43: Consider the vectors $v_1 = (1, 0, -1)$, $v_2 = (1, 1, 0)$, $v_3 = (1, 0, -2) \in \mathbf{R}^3$.

We have $f(v_1) = (3, 0, -5) = v_1 + 2v_3$, $f(v_2) = 2v_2$ and $f(v_3) = (5, 2, -5) = v_1 + 2v_2 + 2v_3$. Therefore the matrix A with respect to the basis $\underline{v} := (v_1, v_2, v_3)$ in the domain and the codomain is

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 2 & 0 & 2 \end{pmatrix}.$$

In order to to compute the matrix B with respect to the standard basis of $\mathbf{R}^3$, we express the standard basis vectors e_1, e_2, e_3 as linear combinations of the v_1, v_2, v_3, then compute $f(e_k)$ using the linearity of f. We have

$$\begin{aligned} e_1 &= 2v_1 - v_3 \\ e_2 &= v_2 - e_1 = -2v_1 + v_2 + v_3 \\ e_3 &= v_1 - v_3. \end{aligned}$$

This implies

$$\begin{aligned} f(e_1) &= 2f(v_1) - f(v_3) = (6, 0, -10) - (5, 2, -5) = (1, -2, -5) \\ f(e_2) &= -2f(v_1) + f(v_2) + f(v_3) = (-6, 0, 10) + (2, 2, 0) + (5, 2, -5) = (1, 4, 5) \\ f(e_3) &= f(v_1) - f(v_3) = (3, 0, -5) - (5, 2, -5) = (-2, -2, 0). \end{aligned}$$ ∎

Therefore

$$B = \begin{pmatrix} 1 & 1 & -2 \\ -2 & 4 & -2 \\ -5 & 5 & 0 \end{pmatrix}.$$

A slightly different way to solve this is to observe that the matrix of f with respect to the basis $\underline{v}$ in the source and the standard basis $\underline{e}$ in the target is $C = \begin{pmatrix} 3 & 2 & 5 \\ 0 & 2 & 2 \\ -5 & 0 & -5 \end{pmatrix}$. We consider the matrix $H = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & -2 \end{pmatrix}$ whose columns are the three vectors v_i, i.e., the base change matrix from the basis $\underline{v}$ to $\underline{e}$. Compute the inverse of this matrix, for example using Theorem 4.80. This gives $H^{-1} = \begin{pmatrix} 2 & -2 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & -1 \end{pmatrix}$. This is the base change matrix from $\underline{e}$

to the basis $\underline{v}$. Then $B = CH^{-1}$: $\mathbf{R}^3_{\underline{e}} \xrightarrow[H^{-1}]{\mathrm{id}} \mathbf{R}^3_{\underline{v}} \xrightarrow[C]{f} \mathbf{R}^3_{\underline{e}}$, which confirms the above computation. Likewise $A = H^{-1}C$, corresponding to $\mathbf{R}^3_{\underline{v}} \xrightarrow[C]{f} \mathbf{R}^3_{\underline{e}} \xrightarrow[H^{-1}]{\mathrm{id}} \mathbf{R}^3_{\underline{v}}$, which confirms the above computation.

The kernel and image of f can be computed by bringing B into row echelon form. The result is $\ker f = \{(x, x, x) | x \in \mathbf{R}\}$, i.e., it is 1-dimensional and has as a basis vector $(1, 1, 1)$. We have $\operatorname{im} f = L((1, 4, 5), (1, -2, -5))$.

For the last part, we form the matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & -2 & 3 \\ -2 & 4 & -2 & t \\ -5 & 5 & 0 & -5 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & -2 & 3 \\ 0 & 6 & -6 & t+6 \\ 0 & 10 & -10 & 10 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & -2 & 3 \\ 0 & 0 & 0 & t \\ 0 & 1 & -1 & 1 \end{array}\right)$$

and bring it into row echelon form as shown above. We find that it has rank 2 (the same as the rank of B), if and only if $t = 0$. and $f^{-1}((3, 0, -5)) = \{(x, x-1, x-2) | x \in \mathbf{R}\}$.

These computations can also be performed with any computer algebra software, such as Wolfram Alpha:

- https://www.wolframalpha.com/input?i=Inverse%28%28%281%2C1%2C1%29%2C%280%2C1%2C0%29%2C%28-1%2C0%2C-2%29%29%29 (compute the inverse of H)

- https://www.wolframalpha.com/input?i=%28%283%2C2%2C5%29%2C%280%2C2%2C2%29%2C%28-5%2C0%2C-5%29%29*Inverse%28%28%281%2C1%2C1%29%2C%280%2C1%2C0%29%2C%28-1%2C0%2C-2%29%29%29 (compute $H^{-1}C$)

- https://www.wolframalpha.com/input?i=kernel%28%281%2C1%2C-2%29%2C%28-2%2C4%2C-2%29%2C%28-5%2C5%2C0%29%29, https://www.wolframalpha.com/input?i=column+space%28%281%2C1%2C-2%29%2C%28-2%2C4%2C-2%29%2C%28-5%2C5%2C0%29%29 (computation of kernel and image of f)

- https://www.wolframalpha.com/input?i=Solve%28%28%281%2C1%2C-2%29%2C%28-2%2C4%2C-2%29%2C%28-5%2C5%2C0%29%29*%28x%2Cy%2Cz%29%3D%283%2Ct%2C-5%29%29 (compute the preimage of a vector)

Solution of Exercise 4.44: We form the matrix of f and bring it to row echelon form:

$$\begin{pmatrix} 1 & -1 & 2 \\ -2 & 3 & -1 \\ 0 & 1 & 3 \\ -1 & 3 & t \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & 3 \\ 0 & 1 & 3 \\ 0 & 2 & t+2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & t-4 \end{pmatrix}$$

The $\dim \operatorname{im} f$ equals the rank of that matrix, which is 2 if $t = 4$ and 3 otherwise.

For no $t \in \mathbf{R}$, f is surjective, since $\dim \operatorname{im} f \leqslant 3$, but it would need to be 4 for f to be surjective. For $t \neq 4$ it is injective. Indeed, the rank-nullity theorem (Theorem 4.26) asserts that $\dim \ker f = 3 - \dim \operatorname{im} f$, which is 0 for $t \neq 4$.

For $t = 4$, we have $\ker f = L(-5, -3, 1)$ and $\operatorname{im} f = L((-1, 3, 1, 3), (1, -2, 0, -1))$.

We have $w \notin \operatorname{im} f$ for all $t \in \mathbf{R}$. This can be shown by forming the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ -2 & 3 & -1 & 1 \\ 0 & 1 & 3 & 0 \\ -1 & 3 & t & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 1 & 3 & 3 \\ 0 & 1 & 3 & 0 \\ -1 & 3 & t & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 0 & 0 & \underline{3} \\ 0 & 1 & 3 & 0 \\ -1 & 3 & t & 1 \end{array}\right)$$

and the underlined entry 3 shows $w \notin \operatorname{im} f$.

For $t = 0$ we have seen above that $\dim \operatorname{im} f = 3$. By Theorem 3.67, a basis v_1, v_2, v_3 of $\operatorname{im} f (\subset \mathbf{R}^4)$ can be extended to a basis of $\mathbf{R}^4$, say v_1, v_2, v_3, v_4. By Proposition 4.39, there is a unique linear map $g : \mathbf{R}^4 \to \mathbf{R}^3$ such that $g(v_i) = e_i$ (for $i = 1, 2, 3$, and e_i denotes the i-th standard basis vector of $\mathbf{R}^3$, and $g(v_4) = 0$ (or any other vector in $\mathbf{R}^3$). Then $g(f(x)) = x$ for any $x \in \mathbf{R}^3$.

C.5 Determinants

Solution of Exercise 5.5: According to Proposition 5.17, the determinant equals $3 \cdot 4 \cdot 5 = 60$.

Solution of Exercise 5.6: Both matrices are non-invertible, since the rows are not linearly independent. Thus they both have determinant 0.

C.6 Eigenvalues and eigenvectors

Solution of Exercise 6.5: The condition $\ker f = L((1,1,1))$ implies that $f(1,1,1) = (0,0,0)$, which we can also rewrite as

$$f((1,1,1)) = 0 \cdot (1,1,1).$$

Thus, this vector is an eigenvector for f, with eigenvalue 0. We therefore have three eigenvectors as follows:

$$\begin{aligned} v_1 &= (1,0,1) \mapsto 2(1,0,1) \\ v_2 &= (2,0,-3) \mapsto -1(2,0,-3) \\ v_3 &= (1,1,1) \mapsto 0(1,1,1). \end{aligned}$$

We check that these three vectors form a basis of $\mathbf{R}^3$ (note that this is therefore an example of an eigenbasis). To this end, we compute the rank of

$$\begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & -3 \\ 1 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -5 \\ 0 & 1 & 0 \end{pmatrix}.$$

This implies that the matrix has rank three, and therefore the three vectors form a basis of $\mathbf{R}^3$. The matrix of f with respect to the basis $\underline{v} = \{v_1, v_2, v_3\}$ is

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In order to compute the matrix of f with respect to the standard basis $\underline{e} = \{e_1, e_2, e_3\}$, we use the usual diagram:

$$\mathbf{R}^3_{\underline{e}} \xrightarrow[K]{\text{id}} \mathbf{R}^3_{\underline{v}} \xrightarrow[\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}]{f} \mathbf{R}^3_{\underline{v}} \xrightarrow[K^{-1}]{\text{id}} \mathbf{R}^3_{\underline{e}}.$$

It turns out that K^{-1} is easier to compute than K. It is given by expressing the v_i in their coordinates in the standard basis vectors, e.g. $v_1 \mapsto \text{id}(v_1) = (1,0,1) = 1e_1 + 0e_2 + 1e_3$. This implies $K^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & -3 & 1 \end{pmatrix}$. We can use this to compute $K =$

$(K^{-1})^{-1}$), cf.(4.71). This inverse (of K^{-1}) can be computed using Theorem 4.80, which gives $K = \frac{1}{5}\begin{pmatrix} 3 & -5 & 2 \\ 1 & 0 & -1 \\ 0 & 5 & 0 \end{pmatrix}$. Then, one computes the product

$$K^{-1}\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}K = \frac{1}{5}\begin{pmatrix} 4 & -10 & 6 \\ 0 & 0 & 0 \\ 9 & -10 & 1 \end{pmatrix}.$$

This is the basis of f with respect to the standard basis.

This is a typical example of the situation that one basis of $\mathbf{R}^3$ may be more adapted to describing a linear map than another one. An eigenbasis, such as v_1, v_2, v_3 gives a particularly simple matrix.

Solution of Exercise 6.7: We have $\det(A_a - t\mathrm{id}_3) = \det\begin{pmatrix} 4-t & 0 & 4 \\ a & 2-t & a \\ -2 & 0 & -2-t \end{pmatrix}$. We compute the determinant by developing the first row (Proposition 5.20), which gives

$$\begin{aligned} \det(A_a - t\mathrm{id}_3) &= (4-t)[(2-t)(-2-t)] + 4[2(2-t)] \\ &= (4-t)(2-t)(-2-t) + 8(2-t) \\ &= (t-2)[(4-t)(2+t) - 8] \\ &= (t-2)[8 + 2t - t^2 - 8] \\ &= (t-2)(-t^2 + 2t) \\ &= -(t-2)^2 t. \end{aligned}$$

The roots of this polynomial, i.e., the eigenvalues are 2 and 0 (regardless of the value of a). The exponent of $t-2$ in the above polynomial is 2, the one for t is 1. This implies that

$$\begin{aligned} 1 \leqslant \dim E_0 \leqslant 1 \text{ for all } t \in \mathbf{R} \\ 1 \leqslant \dim E_2 \leqslant 2 \text{ for all } t \in \mathbf{R}. \end{aligned}$$

According to Method 6.15, A_a will be diagonalizable precisely if $\dim E_2 = 2$. We compute E_2 by bringing $A_a - 2\mathrm{id}$ into reduced row echelon form:

$$\begin{pmatrix} 4-2 & 0 & 4 \\ a & 2-2 & a \\ -2 & 0 & -2-2 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 4 \\ a & 0 & a \\ -2 & 0 & -4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 2 & 0 & 4 \\ a & 0 & a \\ 0 & 0 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ a & 0 & a \\ 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & a-2a \\ 0 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & -a \\ 0 & 0 & 0 \end{pmatrix}.$$

$\blacksquare$

This matrix has rank 1, or equivalently $\dim E_2 = 2$, if and only if $a = 0$. Thus, the matrix A_a is diagonalizable precisely if $a = 0$. The second part of the exercise then has only to be done for $a = 0$, i.e. $A := A_0 = \begin{pmatrix} 4 & 0 & 4 \\ 0 & 2 & 0 \\ -2 & 0 & -2 \end{pmatrix}$. This can be dealt with as in the previous exercises.

Solution of Exercise 6.11: If A and B represent the same map, then $A = PBP^{-1}$ for some invertible matrix P. This implies that $\det A = \det P \det B \det P^{-1} = \det B$. In short, A and B have to have the same determinant. This is true: $\det A = \det B = 9$.

In addition A and B have to have the same characteristic polynomial:

$$\begin{aligned} \chi_A(t) &= \det(A - t\mathrm{id}) \\ &= \det \begin{pmatrix} 1-t & 0 & 2 \\ 0 & 3-t & 0 \\ 0 & 0 & 3-t \end{pmatrix} \\ &= (1-t)(3-t)^2 \\ &= \det(B - t\mathrm{id}) = \chi_B(t). \end{aligned}$$

Again, this is true.

Finally, the dimensions of the eigenspaces of the eigenvalues (1 and 3) need to be equal. For A, the eigenspace $E_{1,A}$ for the eigenvalue $\lambda = 1$ has $\dim E_{1,A} = 2$ (as one computes!). For B instead, $\dim E_{1,B} = 1$. Therefore A and B do *not* represent the same linear map.

Solution of Exercise 6.12: The matrix A has eigenvalues 1 and 2. The eigenspace $E_1 = L(1,0,0)$ and $E_2 = L(1,1,0)$. Their dimensions sum up to 2, which is strictly less than 3, so that A is *not* diagonalizable.

The matrix A^2 therefore has $2^2 = 4$ as an eigenvalue. Since similar matrices have the same eigenvalues, A is *not* similar to A^2.

Solution of Exercise 6.13: We first check that v_1, v_2, v_3 are a basis of $\mathbf{R}^3$. Indeed, one can compute

$$\det \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} = 1 \neq 0,$$

so the rank is 3 and the vectors form a basis.

The condition that v_3 be an eigenvector for the eigenvalue 4 means $f(v_3) = 4v_3 = (4,4,8)$. Acccording to Proposition 4.39, there is a unique linear map f whose value on v_1, v_2, v_3 is prescribed.

We now compute A. We have to express $f(v_i)$ as a linear combination in terms of v_1, v_2, v_3:

$$\begin{aligned} f(v_1) &= (0,0,0) && = 0v_1 + 0v_2 + 0v_3 \\ f(v_2) &= (1,0,3) && = av_1 + bv_2 + cv_3 \\ f(v_3) &= (4,4,8) && = 0v_1 + 0v_2 + 4v_3. \end{aligned}$$

We compute a, b, c above by solving the system

$$(1,0,3) = (a+b+c, b+c, a+b+2c).$$

Thus $b = -c$, $1 = a$ and then $c = 2$. Thus

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 2 & 4 \end{pmatrix}.$$

In order to compute B we could use the base change matrix, but it is also possible to compute B directly. We will express the standard basis vectors $(1,0,0)$ as a linear combination of the v_1, v_2, v_3:

$$\begin{aligned} av_1 + bv_2 + cv_3 &= a(1,0,1) + b(1,1,1) + c(1,1,2) \\ &= (a+b+c, b+c, a+b+2c). \end{aligned}$$

Thus, the equation $(1,0,0) = av_1 + bv_2 + cv_3$ amounts to the linear system

$$\begin{aligned} 1 &= a + b + c \\ 0 &= b + c \\ 0 &= a + b + 2c \end{aligned}$$

One solves this: $a = 1$, $b = 1$, $c = -1$. Similarly, one solves the linear system $(0,1,0) = av_1 + bv_2 + cv_3$. Its solution is $a = -1$, $b = 1$, $c = 0$. Finally, for $(0,0,1) = av_1 + bv_2 + cv_3$ one gets the solution $a = 0$, $b = -1$, $c = 1$.

Thus, since f is linear (!), we have

$$\begin{aligned} f(1,0,0) &= f(v_1 + v_2 - v_3) \\ &= f(v_1) + f(v_2) - f(v_3) \\ &= (0,0,0) + (1,0,3) - 4(1,1,2) \\ &= (-3,-4,-5). \end{aligned}$$

Likewise

$$\begin{aligned} f(0,1,0) &= f(-v_1 + v_2) = -f(v_1) + f(v_2) = (1,0,3) \\ f(0,0,1) &= f(-v_2 + v_3) = -f(v_2) + f(v_3) = (3,4,5) \end{aligned}$$

Therefore (writing $f(1,0,0)$ as the first column etc.), we get

$$B = \begin{pmatrix} -3 & 1 & 3 \\ -4 & 0 & 4 \\ -5 & 3 & 5 \end{pmatrix}.$$

The vector v_t belongs to the image precisely if is a linear combination of the vectors $f(v_1) = 0$, $f(v_2) = (1,0,3)$ and $f(v_3) = (4,4,8)$. This translates into the linear system

$$\begin{aligned} a + 3b &= 2 \\ 4b &= t \\ 3a + 5b &= 5. \end{aligned}$$

One solves the first and third equation to $a = \frac{5}{4}$, $b = \frac{1}{4}$. Therefore, the system has a solution precisely if $t = 1$. (Alternatively, one may also solve the linear system $B\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ t \\ 5 \end{pmatrix}$.)

Solution of Exercise 6.14: The matrix A is invertible precisely iff $\det A = 0$. We compute the determinant, for example using Sarrus' rule, as $\det A = 0 - 24 + 6t - 10t - 0 + 30 = 6 - 4t$. Thus, the condition $\det A = 6 - 4t = 0$ amounts to $t = \frac{3}{2}$. The matrix A is therefore not invertible precisely if $t = \frac{3}{2}$.

We compute the eigenvalues of $A = \begin{pmatrix} 0 & 2 & 2 \\ -3 & -5 & 6 \\ -2 & -2 & 5 \end{pmatrix}$ by computing its characteristic polynomial. It is given by

$$\begin{aligned} \chi_A(t) &= t(5+t)(5-t) - 24 + 12 + 4(-5-t) - 12t + 6(5-t) \\ &= -t^3 + 3t - 2. \end{aligned}$$

One zero of this polynomial is $t = 1$. Dividing the above polynomial by $t - 1$ gives $-t^2 - t + 2$, which has zeroes 1 and -2, respectively. Thus

$$\chi_A(t) = -(t-1)^2(t+2).$$

The eigenvalues of A are therefore $\lambda = 1$ and $\lambda = -2$.

We compute the eigenspaces by bringing $A - \lambda \mathrm{id}$ into row echelon form

$$\begin{aligned} A - (-2)\mathrm{id} = \begin{pmatrix} 2 & 2 & 2 \\ -3 & -3 & 6 \\ -2 & -2 & 3 \end{pmatrix} &\rightsquigarrow \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 2 & 2 & -3 \end{pmatrix} \\ \rightsquigarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} &\rightsquigarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

This matrix has rank 2, and its kernel is thus 1-dimensional. It is spanned by $(1, -1, 0)$. Similarly

$$\begin{aligned} A - \mathrm{id} = \begin{pmatrix} -1 & 2 & 2 \\ -3 & -3 & 6 \\ -2 & -2 & 4 \end{pmatrix} &\rightsquigarrow \begin{pmatrix} 1 & -2 & -2 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ \rightsquigarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

This again has rank 2, so that the eigenspace E_1 is again 1-dimensional. $\blacksquare$ It is spanned by $(2, 0, 1)$.

Since $v = (2, 0, a)$ was requested to be an eigenvector, it will be in one of the two eigenspaces. One sees it must lie in E_1, and $(2, 0, a)$ lies in E_1 precisely if $a = 1$. Thus $v = (2, 0, 1)$, and its eigenvalue is 1.

The matrix A is not diagonalizable, since $\dim E_2 + \dim E_1 = 2 < 3$.

The matrix A is not similar to A^2 since similar matrices have the same determinant. Above we computed $\det A = 6 - 4t$, so for $t = 2$ we have $\det A = -2$, so that $\det A^2 = (-2)^2 = 4 \neq \det A$.

These computations can also be performed using any computer algebra software such as Wolfram Alpha:

- https://www.wolframalpha.com/input?i=Characteristic+polynomial+%28%280%2C2%2C2%29%2C%28-3%2C-5%2C6%29%2C%28-2%2C-2%2C5%29%29 (characteristic polynomial)
- https://www.wolframalpha.com/input?i=Ker+%28+%28%280%2C2%2C2%29%2C%28-3%2C-5%2C6%29%2C%28-2%2C-2%2C5%29%29+-+%28%281%2C0%2C0%29%2C%280%2C1%2C0%29%2C%280%2C0%2C1%29%29%29 (eigenspace)

Solution of Exercise 6.15: The kernel of A is different from $\{0\}$ precisely if A is not invertible or, equivalently, if $\det A = 0$. For example using Sarrus' rule, we have $\det A = 2t - 6$, so $t = 3$.

For $t = 3$, we have $\chi_A(t) = -t^3 + 6t^2 - 8t = -t(t^2 - 6t + 8) = -t(t-2)(t-4)$. The eigenvalues are then 0, 2 and 4.

The eigenspaces are $E_0 = \ker A = L(1, -5, 2)$, $E_2 = \ker(A - 2\mathrm{id}) = L(1, 1, 0)$ and $E_4 = \ker(A - 4\mathrm{id}) = L(-1, 1, 2)$. The matrix $P = \begin{pmatrix} 1 & 1 & -1 \\ -5 & 1 & 1 \\ 2 & 0 & 2 \end{pmatrix}$ (whose columns are the three eigenvectors comprising an eigenbasis) is then such that $PAP^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

If B satisfies $\chi_B(t) = \chi_A(t) = -t(t-2)(t-4)$, its eigenvalues are 0, 2 and 4. These are 3 distinct eigenvalues (i.e., equal to the size of the matrix), so B is diagonalizable.

Solution of Exercise 6.16: We compute $\det A = -4t - 2$, this is 0 precisely if $t = -\frac{1}{2}$. Thus, A is non-invertible for $t = -\frac{1}{2}$ and invertible otherwise.

We compute the characteristic polynomial. It is helpful to develop the determinant of $A - \lambda\mathrm{id}$ along the second column (Proposition 5.20), since there are two 0's in this column, simplifying the formula:

$$\begin{aligned}\det(A - \lambda\mathrm{id}) &= (2-\lambda)\det\begin{pmatrix} 1-\lambda & t \\ 2 & -1-\lambda \end{pmatrix} \\ &= (2-\lambda)(\lambda^2 - 2t - 1).\end{aligned}$$

The eigenvalues of A are therefore $\lambda_1 = 2$ and $\lambda_{2/3} = \pm\sqrt{2t+1}$. The latter two eigenvalues are real numbers precisely if $2t + 1 \geqslant 0$, i.e., if $t \geqslant -\frac{1}{2}$.

An eigenvalue appears with multiplicity 2 in the above characteristic polynomial precisely if either $2t + 1 = 0$ (so that $\lambda_2 = \lambda_3$), i.e., $t = -\frac{1}{2}$, or if $\sqrt{2t+1} = 2$, i.e., if $t = \frac{3}{2}$.

For $t = -\frac{1}{2}$, we compute the eigenspaces for the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 0$. These are both 1-dimensional, respectively given by $E_2 = L(0, 1, 0)$ and $E_0 = L(2, -3, 4)$. The sum of their dimensions is 2, which is less than 3, so A is not diagonalizable (Method 6.15).

C.7 Euclidean spaces

Solution of Exercise 7.2: The given equation describing U can be rewritten as

$$\begin{pmatrix} 1 & -1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \left\langle \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle = 0,$$

which gives, with free parameters $y = a$ and $z = b$, $x = a - 3b$. Thus $U = \{(a - 3b, a, b) \mid a, b \in \mathbf{R}\} = L((1, 1, 0), (-3, 0, 1))$.

As for the orthogonal complement $U^\perp$, we know that $\dim U^\perp = 3 - \dim U = 1$ (Corollary 7.30). In order to compute $U^\perp$ we need to find the vectors that are orthogonal to U. The above equation tells us that U is the orthogonal complement of the vector $\begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$.

Therefore $U^\perp = L(\begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix})$.

In order to compute the projection of t onto U, we apply the Gram–Schmidt orthogonalization method. The vector $v_1 = (1,1,0)$ has norm $\|v_1\| = 2$, so that $w_1 = \frac{1}{\sqrt{2}}(1,1,0)$. Then

$$\begin{aligned} w_2' := v_2 - \langle v_2, w_1\rangle w_1 &= (-3,0,1) - \left[(-3,0,1)\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}\right] w_1 \\ &= (-3,0,1) + \frac{3}{\sqrt{2}}\frac{1}{\sqrt{2}}(1,1,0) \\ &= (-3,0,1) + (\frac{3}{2}, \frac{3}{2}, 0) \\ &= (-\frac{3}{2}, \frac{3}{2}, 1). \end{aligned}$$

We finally normalize this vector: $\|w_2'\| = \sqrt{\frac{9}{4} + \frac{9}{4} + 1} = \sqrt{\frac{11}{2}}$. Therefore

$$w_2 = \sqrt{\frac{2}{11}}(-\frac{3}{2}, \frac{3}{2}, 1) = (-\frac{3}{\sqrt{22}}, \frac{3}{\sqrt{22}}, \sqrt{\frac{2}{11}}).$$

According to Theorem 7.24, the orthogonal projection is given by

$$\begin{aligned} \langle t, w_1\rangle w_1 + \langle t, w_2\rangle w_2 &= \left[(0,1,5)\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}\right] w_1 + \left[(0,1,5)\sqrt{\frac{2}{11}}\begin{pmatrix} -\frac{3}{2} \\ \frac{3}{2} \\ 1 \end{pmatrix}\right] w_2 \\ &= \frac{1}{\sqrt{2}} w_1 + \sqrt{\frac{2}{11}}\frac{13}{2} w_2 \\ &= (\frac{1}{2}, \frac{1}{2}, 0) + \frac{13}{11}(-\frac{3}{2}, \frac{3}{2}, 1) \\ &= (\frac{-28}{22}, \frac{50}{22}, \frac{13}{11}) \\ &= (-\frac{14}{11}, \frac{25}{11}, \frac{13}{11}). \end{aligned}$$

∎

An alternative way to solve this, with slightly fewer computations, is the following: since $t = t_U + t_\perp$ is the unique decomposition, we can compute $t_U = t - t_\perp$. By the positive definitness of $\langle -, - \rangle$ we have $(U^\perp)^\perp = U$, so we can also apply Gram–Schmidt to $U^\perp$. This is somewhat simpler since $U^\perp$ has dimension 1. Applying Gram–Schmidt to $v = (1,-1,3)$, we have $\|v\| = \sqrt{11}$, so that $w = \frac{1}{\sqrt{11}}(1,-1,3)$ is an ortho*normal* basis of $U^\perp$. Then, again by

Theorem 7.24

$$\begin{aligned} t_\perp &= \langle t, w\rangle w \\ &= \left[(0,1,5)\frac{1}{\sqrt{11}}(1,-1,3)\right] w \\ &= \frac{14}{\sqrt{11}}\frac{1}{\sqrt{11}}(1,-1,3) \\ &= (\frac{14}{11}, -\frac{14}{11}, \frac{42}{11}). \end{aligned}$$

Therefore

$$t_U = t - t_\perp = (0,1,5) - (\frac{14}{11}, -\frac{14}{11}, \frac{42}{11}) = (-\frac{14}{11}, \frac{25}{11}, \frac{13}{11}).$$

Solution of Exercise 7.4: U is given by the solutions of the homogeneous linear system

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The left hand matrix can be brought to reduced row echelon form as $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, so that $z = a$ is a free parameter and $x = 0$, $y = -a$. This shows that $U = \{(0,-a,a) \mid a \in \mathbf{R}\}$ and $(0,-1,1)$ is a basis vector of U. The orthogonal complement consists of vectors orthogonal to $(0,-1,1)$. As in the solution of Exercise 7.2 above, U has been defined as the orthogonal complement of the two vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. These vectors therefore constitute a basis of $U^\perp$.

We compute the orthogonal projection of $t = (5,1,3)$ onto U and $U^\perp$. This can be done using Gram–Schmidt orthogonalization as above, but also by solving the linear system

$$\begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix} = a\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + b\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

This is quickly solved as $\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$, so that the projection of t onto U is $a \cdot \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$.

Solution of Exercise 7.5: The orthogonal complement of T is given by vectors (x, y, z) such that $x - 3z = 0$. I.e., with free parameters $y = a$, $z = b$, $x = 3b$. Thus

$$T^{\perp} = L((3, 0, 1), (0, 1, 0)).$$

Solution of Exercise 7.6:

We want to find U such that $U \oplus U^{\perp} = \mathbf{R}^3$ and $t = (1, 5, 6) + t_{\perp}$, with $(1, 5, 6) \in U$ and $t_{\perp} \in U^{\perp}$. This in particular means that $(1, 5, 6) \perp t_{\perp}$. Solving the equation

$$(1, 1, 0) = (1, 5, 6) + t_{\perp}$$

gives $t_{\perp} = (1, 1, 0) - (1, 5, 6) = (0, -4, -6)$. But

$$\langle (1, 5, 6), (0, -4, -6) \rangle = -20 + 36 = 16 \neq 0,$$

so these two vectors are *not* orthogonal. Therefore there is *no* such subspace U.

We solve the second part similarly: we have $t - (1, 1, 1) = (1, -1, 0)$, so the Ansatz is $t_{\perp} = (1, -1, 0)$. We compute $\langle (1, 1, 1), (1, -1, 0) \rangle = 0$, so these vectors are orthogonal. We now compute U: $(1, 1, 1) \in U$, so that we can take $U = L(1, 1, 1)$.

Solution of Exercise 7.7: According to Proposition 7.44, the unique point $v \in \mathbf{R}^3$ that is lying on L and as close as possible to the origin is given by

$$v = v_0 - p_L(v_0),$$

where $v_0 = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$. We will use Theorem 7.24 in order to compute this. The underlying subspace W of L is spanned by $v_1 = \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}$. Renormalizing this vector to norm one, gives $w_1 = \frac{v_1}{\|v_1\|} = \frac{v_1}{\sqrt{18}}$. This vector w_1 is therefore an orthonormal basis of W. We then have

$$p_L(v_0) = \langle v_0, w_1 \rangle w_1 = \frac{24}{18} \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}$$

and therefore

$$v = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} - \frac{24}{18}\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} -1/3 \\ 5/3 \\ -1/3 \end{pmatrix}.$$

This vector has norm

$$\|v\| = \sqrt{27/9} = \sqrt{3}.$$

Solution of Exercise 7.8: The subspace W underlying L is spanned by $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, while the equatios for L' are equivalent to $x = z$ and $y = z + 2$. Therefore, this line has the underlying subspace $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ as well. Therefore, the lines are parallel. Let $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ be the direction vector for both lines. Below, we will consider its renormalization to norm 1, which is $w_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Let $v_0 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ and $v_0' = \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix}$. Then the distance vector $d = v_0 - v_0' = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ is not orthogonal to L ($\langle d, w\rangle \neq 0$). We compute the orthogonal projection of d onto $W^\perp$ by computing

$$d - \langle d, w_1\rangle w_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\langle d, w\rangle = \frac{1}{3}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

This vector has norm $\sqrt{2/3}$, which is therefore the distance of L and L'.

Solution of Exercise 7.9: The two lines have underlying vector spaces $W = L(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix})$ and $W' = L(\begin{pmatrix} 1 \\ -1 \\ 1/2 \end{pmatrix})$ respectively. These two one-dimensional subspaces are not contained in each other: the two vectors are linearly independent.

Let $d = v_0 - v_0' = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$. We compute the orthogonal complement of

$$Z = W + W' = L(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} (\\ 1 \\ -1 \end{pmatrix} 1/2) = L(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} (\\ 2 \\ -2 \end{pmatrix} 1).$$

It is given by the equations

$$\begin{aligned} x + y + z &= 0 \\ 2x - 2y + z &= 0 \end{aligned}$$

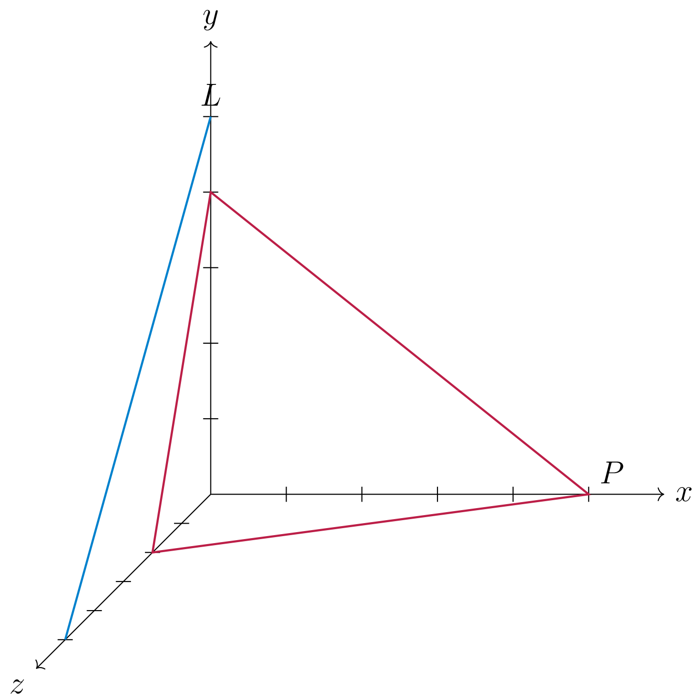
which can be solved as $L(\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix})$. This vector has norm 6, and is renormalized to norm 1 as $w_\perp = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$. We compute the orthogonal projection of d onto $Z^\perp$ as

$$p_{Z^\perp}(d) = d - \langle d, w_\perp \rangle w_\perp = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} - \frac{-1}{6}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} -11 \\ 7 \\ -2 \end{pmatrix}.$$

This vector has norm $\sqrt{\frac{29}{6}}$.

In particular, this distance is positive. This, together with the above observation that $W \neq W'$ means that the lines are skew. (If one is only required to show that the lines are skew, without computing their distance, one can also solve the linear system given by the intersection of L and L'; one finds out that this system has no solution, so the lines are skew.)

Solution of Exercise 7.10: A useful way to sketch lines and planes is by considering some points on them where several coordinates are zero. In the case of P, three such points are $(5, 0, 0), (0, 4, 0), (0, 0, 2)$, while for L two such points are $(0, 5, 0), (0, 0, 5)$. This leads to the following sketch:



The equation can be rewritten as

$$\left\langle \begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle = 20.$$

Thus, the underlying vector space W is the orthogonal complement of the vector $\begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}$.

The line L has as its underlying vector space $W' = L(\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix})$. If P and L are parallel, then $W' \subset W$. This means that $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}$. Their scalar product is $5 - 10 = -5 \neq 0$, so that these vectors are *not* orthogonal and therefore L and P are not parallel.

We compute the distance of P to the origin using Proposition 7.46:

$$d(0, P) = \frac{20}{\|\begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}\|} = \frac{20}{\sqrt{141}}.$$

We compute the closest point by determining the (unique) intersection point $P \cap W^\perp$. We are thus seeking the real number r such that $r\begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 4r \\ 5r \\ 10r \end{pmatrix}$ lies on the plane P, i.e., such that

$$\left\langle \begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 4r \\ 5r \\ 10r \end{pmatrix} \right\rangle = 20.$$

The left hand side equals $r(16 + 25 + 100) = 141r$, so that $r = \frac{20}{141}$, and therefore the point in P that is as close as possible to the origin is

$$\frac{20}{141}\begin{pmatrix} 4 \\ 5 \\ 10 \end{pmatrix}.$$

Solution of Exercise 7.11: We apply Theorem 7.37, according to which a matrix is *orthogonally* diagonalizable if and only if it is symmetric. This excludes $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$. The other two matrices are symmetric and therefore orthogonally diagonalizable. The matrix $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ is already a diagonal matrix, so for $P = \mathrm{id}$ the matrix PAP^{-1} is diagonal. The standard basis vectors e_1, e_2 are an orthonormal eigenbasis.

For $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, we compute the eigenvalues as was indicated after Theorem 7.37.

$$\begin{aligned} \lambda_{1/2} &= \frac{a+d}{2} \pm \sqrt{\frac{(a-d)}{4} + b^2} \\ &= 1 \pm 2. \end{aligned}$$

Thus the eigenvalues are -1 and 3. We compute the eigenspaces. The matrix $A-(-1)\mathrm{id} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$ has kernel given by $w_1 := e_1 - e_2$. The matrix $A - 3\mathrm{id} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$ has kernel given by $w_2 := e_1 + e_2$. These vectors are orthogonal:

$$\langle e_1 - e_2, e_1 + e_2 \rangle = \left\langle \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = 0.$$

However, they are not normal, so an orthonormal eigenbasis for A is given by

$$\frac{w_1}{\|w_1\|} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } \frac{w_2}{\|w_2\|} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Solution of Exercise 7.12: If $Av = \lambda v$, $Aw = \lambda w$, we compute

$$\begin{aligned} \lambda\langle v, w\rangle &= \langle \lambda v, w\rangle \\ &= \langle Av, w\rangle \\ &= (Av)^T w \\ &= v^T A^T w \\ &= v^T A w && \text{since } A = A^T \\ &= \langle v, Aw\rangle \\ &= \langle v, \mu w\rangle \\ &= \mu\langle v, w\rangle. \end{aligned}$$

Since $\lambda \neq \mu$, this forces $\langle v, w\rangle = 0$.

Solution of Exercise 7.13: By definition, the hyperplane P is given by the equation

$$\left\langle (2,0,1,-1)^T, (x_1, \dots, x_4)^T \right\rangle = 4.$$

In other words, the underlying sub-vector space $W \subset \mathbf{R}^4$ of that hyperplane is the subspace

$$\left\langle (2,0,1,-1)^T, (x_1, \dots, x_4)^T \right\rangle = 0,$$

or, what is the same, the orthogonal complement of $(2,0,1,-1)^T$.

The underlying subspace W'_t of the line L_t is spanned by $(t,1,0,-1)$. By Definition 7.41, P is parallel to L_t if $W \subset W'_t$ (which is impossible for dimension reasons) or if $W'_t \subset W$. The latter is equivalent to $(t,1,0,-1) \in W$ or, yet equivalently,

$$\left\langle (t,1,0,-1)^T, (2,0,1,-1)^T \right\rangle = 0.$$

This equates to $2t + 1 = 0$ or $t = -\frac{1}{2}$.

We now compute the pair(s) of points realizing the minimal distance between P and $L := L_{-\frac{1}{2}}$. We will use that these are exactly the points (p, l) such that $p - l \perp W$ and $p - l \perp W'$.

The point $l \in L = L_{-\frac{1}{2}}$ is of the form

$$\begin{aligned} l &= (1,0,0,1) + r(-\frac{1}{2}, 1, 0, -1) \\ &= (1-s, 2s, 0, 1-2s) \qquad \text{with } s = \frac{r}{2}. \end{aligned}$$

On the other hand a point $p = (x_1, \ldots, x_4) \in P$ satisfies

$$2x_1 + x_3 - x_4 = 4,$$

so we can take $x_1 = a$, $x_2 = b$ and $x_3 = c$ as free parameters and get $x_4 = 2a + c - 4$. That is $p = (a, b, c, 2a + c - 4)$. Therefore

$$p - l = (a + s - 1, b - 2s, c, 2a + c + 2s - 5).$$

Note that a, b, c, s are the unknowns. We now determine the values of these unknowns. The orthogonal complement of W, $W^\perp = (((2,0,1,-1)^T)^\perp)^\perp = L(2,0,1,-1)$. Here we use that for a subspace $U \subset \mathbf{R}^n$, we have

$$(U^\perp)^\perp = U.$$

(Indeed, $U \subset (U^\perp)^\perp$: for $u \in U$ and $u' \in U^\perp$ we have $\langle u, u' \rangle = 0$, so that $u \in (U^\perp)^\perp$. Both vector subspaces of $\mathbf{R}^n$ have the same dimension, and therefore they agree.)

This means that we are looking for $a, b, c, s \in \mathbf{R}$ such that the following conditions are satisfied

- $p - l$ is a multiple of $(2,0,1,-1)$, say $\lambda(2,0,1,-1) = (2\lambda, 0, \lambda, -\lambda)$. This translates into a linear system

$$\begin{aligned} a + s - 1 &= 2\lambda \\ b - 2s &= 0 \\ c &= \lambda \\ 2a + c + 2s - 5 &= -\lambda. \end{aligned}$$

- We now spell out the condition that $p - l$ is orthogonal to $(-\frac{1}{2}, 1, 0, -1)^T$ or equivalently, to $(-1, 2, 0, -2)$:

$$\begin{aligned} 0 &= \left\langle p - l, (-1,2,0,-2)^T \right\rangle \\ &= -(a + s - 1) + 2(b - 2s) - 2(2a + c + 2s - 5) \\ &= -5a + 2b - 2c - 9s + 11. \end{aligned}$$

We obtain a linear system in the 5 unknowns a, b, c, s, λ:

$$\begin{aligned} a + s - 2\lambda &= 1 \\ b - 2s &= 0 \\ c - \lambda &= 0 \\ 2a + c + 2s + \lambda &= 5 \\ -5a + 2b - 2c - 9s &= -11. \end{aligned}$$

One solves this using Gaussian eliminiation (or, using an appropriate computer algebra system, cf. https://www.wolframalpha.com/input?i=Solve%28%28-5a%2B2b-2c-9s%2B11%3D0%2Ca%2Bs-1%3D2l%2Cb-2s%3D0%2Cc%3Dl%2C2a%2Bc%2B2s-5%3D-l%29%29) and obtains as solutions the vectors

$$(a, b = 4 - 2a, c = \frac{1}{2}, \lambda = \frac{1}{2}, s = 2 - a).$$

Thus

$$\begin{aligned} p &= (a, 4 - 2a, \frac{1}{2}, 2a - \frac{7}{2}) \\ l &= (1 - (2 - a), 2(2 - a), 0, 1 - 2(2 - a)) = (a - 1, 4 - 2a, 0, -3 + 2a). \end{aligned}$$

This implies that for *any* point $l \in L$, there is a unique point $p \in P$ such that (p, l) realizes the minimal distance between P and L.

Solution of Exercise 7.17: We compute in fact directly an ortho*normal* basis, which in particular is then orthogonal. (The property of also being normal is convenient further below.) We apply the Gram–Schmidt procedure:

$$\begin{aligned} w_1 &= \frac{1}{\sqrt{6}}(1, 2, 0, -1), \\ w_2' &= (0, -4, 3, 4) - \left\langle (0, -4, 3, 4), \frac{1}{\sqrt{6}}(1, 2, 0, -1) \right\rangle \frac{1}{\sqrt{6}}(1, 2, 0, -1) \\ &= (0, -4, 3, 4) + 2(1, 2, 0, -1) \\ &= (2, 0, 3, 2), \\ w_2 &= \frac{1}{\sqrt{17}}(2, 0, 3, 2). \end{aligned}$$

The vectors w_1 and w_2 then constitute an orthonormal basis of U.

We compute the orthogonal complement of U by solving the equations (for some $v \in \mathbf{R}^4$)

$$\begin{aligned} \langle v, (1,2,0,-1)\rangle &= 0 \\ \langle v, (0,-4,3,4\rangle &= 0. \end{aligned}$$

We consider the matrix of the resulting linear system, and bring it into row echelon form:

$$\begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & -4 & 3 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 1 & -\frac{3}{4} & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & \frac{3}{2} & 1 \\ 0 & 1 & -\frac{3}{4} & 1 \end{pmatrix}.$$

Thus, if $v = (x_1, \ldots, x_4)$, then x_3 and x_4 are free variables, so that a basis of $U^\perp$ is given by the two vectors $(-\frac{3}{2}, \frac{3}{4}, 1, 0)$ and $(-1, 1, 0, 1)$.

We compute the orthogonal projection

$$\begin{aligned} p_U(v) &= \langle v, w_1\rangle w_1 + \langle v, w_2\rangle w_2 \\ &= (3,2,3,1). \end{aligned}$$

If we write $w = l + l^\perp$ with $l^\perp \in L^\perp$, then $l^\perp = w - l = (2, -, 1, 0, 2) - (1,1,2,0) = (1,-2,-2,2)$. This vector would need to be orthogonal to $(1,1,2,0)$, which however it is not (since their scalar product is $-5 \neq 0$). Therefore, such a subspace L does not exist.

These computations may be carried out also using a computer algebra system, for example the Gram–Schmidt orthogonalization procedure like so: https://www.wolframalpha.com/input?i=Orthogonalize%5B%7B%7B1%2C2%2C0%2C-1%7D%2C%7B0%2C-4%2C3%2C4%7D%7D%5D.

Solution of Exercise 7.18: We first compute L in the form $L = v_0 + W$. We choose $v_0 = (0,1,-1)$. In addition, the point $(1,2,1)$ is also lying in L. Therefore W is spanned by $(1,2,1) - (0,1,-1) = (1,1,2)$.

Similarly, we can compute $M = (1,-2,0) + L((2,1,1))$. The two vectors $(1,1,2)$ and $(2,1,1)$ are linearly independent, so that L and M are not parallel nor identical. We compute the intersection $L \cap M$. We have the equations $1 - y = x = 2y + 1$, so that $y = 0$. We also have $2x - 1 = z = y + 2$, so that $z = 2$ and $x = \frac{3}{2}$, but then we get a contradiction to $1 - y = x$. Thus $L \cap M = \varnothing$. Therefore, L is skew to M.

We compute the requested plane P by observing that its underlying vector space W is spanned by $(1,-1,2)$ and $(2,1,1)$. Moreover, the point $(1,0,2) \in P$. Thus $P = (1,0,2) + L((1,-1,2),(2,1,1))$. The orthogonal complement of W is spanned by $(1,-1,-1)$, as one sees by solving the linear system $a-b+2c=0, 2a+b+c=0$. Thus, P is of the form

$$\langle v,(1,-1,-1)\rangle = d.$$

To compute the number d, we use that $(1,0,2) \in P$, so that $d = \langle (1,0,2),(1,-1,-1)\rangle = -1$.

A general point on M is of the form $(1+2r, r, 2+r)$. With $l = (0,1,-1)$, the vector $v := m - l = (1+2r, r-1, 3+r)$. The line spanned by this vector is parallel to the plane defined by the equation $3x - z = 0$ precisely if $\langle v,(3,0,-1)\rangle = 0$. This leads to the equation

$$3 + 6r - 3 - r = 0,$$

i.e., $r = 0$, so that $m = (1,0,2)$.

We compute the coordinates of $r_\alpha = (x,y,z)$ as $z = \alpha$, $x = \frac{1}{2}(1+\alpha)$, $y = \frac{1}{2}(1-\alpha)$. Similarly, $s_\alpha = (2\alpha-3, \alpha-2, \alpha)$. The midpoints $m_\alpha = \frac{r_\alpha + s_\alpha}{2} = (\frac{5\alpha-5}{4}, \frac{\alpha-3}{4}, \alpha)$ are precisely the points on the line

$$(-\frac{5}{4}, -\frac{3}{4}, 0) + L((\frac{5}{4}, \frac{1}{4}, 1)).$$

Solution of Exercise 7.19: We have $L = p + L(q-p)$, i.e., $L = (3,1,0) + L(-3,0,3)$. (Other solutions are possible as well, e.g., $L = q + L(-3,0,3)$.) We determine whether r lies on L by considering the linear system

$$\begin{aligned} 3 - 3t &= -3 \\ 1 &= 0 \\ 3t &= -3. \end{aligned}$$

The second equation is a contradiction, therefore there is no $t \in \mathbf{R}$ satisfying this linear system. Thus $r \notin L$.

We describe the plane P containing the points p, q, r in two ways:

$$P = p + L(q-p, r-p) = (3,1,0) + L((-3,0,3),(-6,-1,-3)).$$

The orthogonal complement of $w_1 = (-3,0,3)$ and $w_2 = (-6,-1,-3)$ is the line spanned by $a := (1,-9,1)$. Thus

$$P = \{x \in \mathbf{R}^3 | \langle x,a\rangle = \langle p,a\rangle\}$$

which we can write out as

$$P = \{x \in \mathbf{R}^3 | \left\langle x, \begin{pmatrix} 1 \\ -9 \\ 1 \end{pmatrix} \right\rangle = -6\}$$

or, equivalently

$$P = \{x = (x_1, x_2, x_3) | x_1 - 9x_2 + x_3 = -6\}.$$

Solution of Exercise 7.20: The line M, given to us by the systen $x + z = 2$, $x - 2y = 2$ is also described as $M = (2, 0, 0) + L(2, 1, -2)$. Its underlying vector space is therefore spanned by $(2, 1, -2)$, while the underlying vector space of L is spanned by $(1, 0, -1)$. These two vectors are linearly independent. We compute the intersection of L and M by taking a general point $l = (3 + t, 1, -t) \in L$ and $m = (2 + 2s, s, -2s) \in M$, with $t, s \in \mathbf{R}$. The three coordinates of the equation $l = m$ read:

$$\begin{aligned} 3 + t &= 2 + 2s \\ 1 &= s \\ -t &= -2s. \end{aligned}$$

The second and third equations give $s = 1$, $t = 2$. The first then gives $5 = 4$, which is a contradiction. Thus, there is no point lying on L and on M, i.e., $L \cap M = \emptyset$. Thus L and M are skew, so no plane contains L and M.

Solution of Exercise 7.21: Writing $q = (x, y, z)$ for $x, y, z \in \mathbf{R}$, the line M is then given by $M = p + L(q - p) = (-1, -1, -1) + L(x + 1, y + 1, z + 1)$.

The line M will be orthogonal to L precisely if

$$(x + 1, y + 1, z + 1) \perp (1, 0, -1),$$

i.e., if $x + 1 - z - 1 = 0$, i.e., if $x = z$.

As in the solution of Exercise 7.20 above, the line M (with $x = z$) intersects L precisely if the linear system

$$\begin{aligned} 3 + t &= -1 + s(x + 1) \\ 1 &= -1 + s(y + 1) \\ -t &= -1 + s(x + 1) \end{aligned}$$

has a solution $(s,t) \in \mathbf{R}^2$. To get a clearer view of this, we write down the augmented matrix for this linear system, keeping in mind that x, y are parameters in this linear system, and s, t are the unknowns. Note this is a linear system in two unknowns, but three equations.

$$\left(\begin{array}{cc|c} -x-1 & 1 & -4 \\ -y-1 & 0 & -2 \\ -x-1 & -1 & -1 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} -x-1 & 1 & -4 \\ -y-1 & 0 & -2 \\ -2x-2 & 0 & -5 \end{array}\right)$$

$$\overset{\text{if } y \neq -1}{\rightsquigarrow} \left(\begin{array}{cc|c} -x-1 & 1 & -4 \\ 1 & 0 & \frac{2}{y+1} \\ -2x-2 & 0 & -5 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{cc|c} 0 & 1 & -4+2\frac{x+1}{y+1} \\ 1 & 0 & \frac{2}{y+1} \\ 0 & 0 & -5+2\frac{2x+2}{y+1} \end{array}\right).$$

If $y = -1$, then we have no solution in the above system, due to the row $0\ 0 \mid -2$ in this case. Otherwise, for $y \neq -1$, we can perform the elementary row operations as indicated above. This system has *no* solution if the term $-5 + 2\frac{2x+2}{y+1} \neq 0$. If, on the contrary, the bottom right entry is zero, then the system does have a unique solution (since we then have two leading ones in the matrix). This bottom right entry is zero precisely if $5(y+1) = 4x+4$, or, equivalently, if $y = \frac{4x-1}{5}$.

We therefore find that the line M through p and $q = (x, y, z)$ intersects L orthogonally precisely if $z = x$, $y = \frac{4x-1}{5}$ and $y \neq -1$.

Solution of Exercise 7.22: Any line M passing through p is of the form $M = p + L(w)$. We compute $w = (a, b, c)$ by considering the two conditions on M:

- $M \perp L$ holds exactly if $w \perp (1,1,1)$, i.e., if $a + b + c = 0$.
- $M \subset P$ holds exactly if $w \perp (3,-4,1)$. Indeed, the underlying subspace of P is given by $3x - 4y + z = 0$, i.e., it is the orthogonal complement of $(3,-4,1)$. That is, $3a - 4b + c = 0$.

We have to solve the homogeneous linear system associated to the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 3 & -4 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & -7 & -2 \end{pmatrix}$$

so a solution is given by $b = 2$, $c = -7$ and $a = 5$. (Any other non-zero multiple of $(5, 2, -7)$ is also a solution.) Therefore

$$M = p + L(w) = (0, 1, 6) + L(5, 2, -7).$$

Solution of Exercise 7.23: The cartesian equations of the line L_2 are obtained by inserting $x = t$ into the equation for z, which gives

$$\begin{aligned} y &= 2 \\ x + z &= 4. \end{aligned}$$

We compute the intersection of L_1 and L_2, and at the same time determine whether they are parallel or not. The linear system for $L_1 \cap L_2$ is (the first two equations are for L_2, the latter two for L_1)

$$\begin{aligned} y &= 2 \\ x + z &= 4 \\ 2x - y &= -3 \\ y + z &= -2. \end{aligned}$$

We form the augmented matrix of this linear system and bring it into row echelon form:

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 2 & -1 & 0 & -3 \\ 0 & 1 & 1 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & -1 & -2 & -11 \\ 0 & 1 & 1 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & -2 & -9 \\ 0 & 0 & 1 & -4 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & -2 & -9 \\ 0 & 0 & 0 & -17 \end{array}\right).$$

If we only consider the first three columns (corresponding to the three variables), the matrix has rank 3. This shows that the lines are not parallel (otherwise the rank would be 2). The full matrix has rank 4. Therefore there is no solution of the linear system and the lines are skew.

Part (2): The plane π contains L_2 and the direction of L_1 (i.e., the underlying sub-vector space) is contained in the directions of π

(i.e., its underlying sub-vector space). We therefore first compute the direction of L_1. The system

$$\begin{aligned} L_1 : 2x - y &= -3 \\ y + z &= -2 \end{aligned}$$

leads to $y = 2x + 3$, $z = -y - 2 = -2x - 5$, in which x is a free variable. Thus

$$\begin{aligned} L_1 &= \{(x, 2x+3, -2x-5) | x \in \mathbf{R}\} \\ &= \{(0, 3, -5) + x(1, 2, -2) | x \in \mathbf{R}\} \end{aligned} \tag{C.4}$$

Thus the direction of L_1 is $(1, 2, -2)$. (Alternatively, one may observe that the points $p = (0, 3, -5)$ and $q = (1, 5, -7)$ satisfy the linear system describing L_1, so that $L_1 = (0, 3, -5) + L(q - p)$.)

The plane π is of the form $\pi = \{v = (x, y, z) \in \mathbf{R}^3 | \langle v, a\rangle = d\}$ for some vector $a = (a_1, a_2, a_3)$ and $d \in \mathbf{R}$. I.e., $\pi = \{(x, y, z) | a_1 x + a_2 y + a_3 z = d\}$. Its underlying subvector space is $a_1 x + a_2 + a_3 z = 0$. The condition L_1 being parallel to π translates into $\langle a, v_1\rangle = 0$, where $v_1 = (1, 2, -2)$ is the direction of L_1. The condition $L_2 \subset \pi$ translates into $\langle a, v_2\rangle = 0$, where $v_2 = (1, 0, -1)$ is the direction vector of L_2. These two conditions give the linear system

$$\begin{aligned} a_1 + 2a_2 - 2a_3 &= 0 \\ a_1 - a_3 &= 0. \end{aligned}$$

Thus $a_1 = a_3$ and $a_1 = 2a_2$, and a_2 (say) is a free variable. We choose $a_2 = 1$ (note: any other non-zero number a_2 would eventually give rise to the same plane π below), so that $a = (a_1, a_2, a_3) = (2, 1, 2)$. Thus, $\pi = \{(x, y, z) | 2x + y + 2z = d\}$, where we still need to compute d. It is enough to choose d such that π contains any point of L_2, for example $e := (0, 2, 4)$. We compute $\langle a, e\rangle = 2 \cdot 0 + 1 \cdot 2 + 2 \cdot 4 = 10$. Thus, $d = 10$, so that

$$\pi = \{(x, y, z) | 2x + y + 2z = 10\}.$$

Part (3): By construction L_1 is parallel to π, so to compute the distance of π and L_1, we may choose any point $q \in L_1$ and to compute its orthogonal projection onto π, which we call r. Then $d(\pi, L_1) = d(q, r)$.

The line passing through q and r is orthogonal to π, so it has direction vector $(2, 1, 2)$. We choose $q = (0, 3, -5)$. The line passing through q and orthogonal to π is then of the form

$$M = (0,3,-5) + L(2,1,2),$$

its points are of the form $r_t = (2t, 3+t, -5+2t)$ where $t \in \mathbf{R}$. The intersection of that line with π is given by the unique value of t such that $r_t \in \pi$, i.e.,

$$2(2t) + (3+t) + 2(-5-2t) = 10.$$

This simplifies to $9t = 17$ or $t = \frac{17}{9}$. Thus $r = r_{17/9} = (\frac{34}{9}, \frac{44}{9}, -\frac{11}{9})$. We then have

$$\begin{aligned} d(L_1, \pi) = d(q, r) &= \|(0,3,-5) - (\frac{34}{9}, \frac{44}{9}, -\frac{11}{9})\| \\ &= \|(-\frac{34}{9}, -\frac{17}{9}, -\frac{34}{9})\| = \frac{17}{3}. \end{aligned}$$

Part (4): we already know the lines are skew, so there will be unique points in L_1 and L_2, respectively, realizing the minimal distance. We compute these points using Theorem 7.56(4). A general point $p_t \in L_2$ is of the form $p_t = (t, 2, 4-t)$ (with $t \in \mathbf{R}$). By (C.4), a general point in L_1 is of the form $q_s = (s, 3+2s, -5-2s)$ (with $s \in \mathbf{R}$). The vector $p_t - q_s = (s-t, 1+2s, -9-2s+t)$ needs to be orthogonal to the directions of L_1 and L_2, which we computed above as $(1,2,-2)$ and $(1,0,-1)$. This leads to the linear system

$$\begin{aligned} 0 &= s - t + 2(1+2s) - 2(-9-2s+t) = 9s - 3t + 20 \\ 0 &= s - t - (-9-2s+t) = 3s - 2t + 9 \end{aligned}$$

This can be solved to $t = \frac{7}{3}$ and $s = -\frac{13}{9}$, which gives

$$\begin{aligned} p_{7/3} &= (\frac{7}{3}, 2, \frac{5}{3}), \\ q_{-13/9} &= (-\frac{13}{9}, -\frac{17}{9}, -\frac{19}{9}). \end{aligned}$$

Solution of Exercise 7.24: We first compute a basis of U by computing the kernel of the matrix

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & -1 \end{pmatrix}$$

The third and fourth variables are free, and a basis of U is given by the vectors $u_1 = (0, 1, 0, 1)$ and $u_2 = (-1, 0, 1, 0)$. These two vectors happen to be orthogonal, so they form an orthogonal basis. (If one chooses a different basis of U, one may apply the Gram–Schmidt algorithm to make them orthonormal, for example.)

Since $\dim U^\perp = 4 - \dim U = 2$, it suffices to find any non-zero vector $w_2 \in U^\perp$ that is also orthogonal to w_1. We are thus looking for a non-zero vector in $L((-1, 0, 1, 0), (0, 1, 0, 1), (1, 1, -1, -1))^\perp$. This is the kernel of the matrix

$$\begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix},$$

so if $w_2 = (x_1, \dots, x_4)$, we obtain $x_2 = x_4 = 0$ and $x_1 = x_3$, so $w_2 = (1, 0, 1, 0)$ satisfies the requested properties.

The orthogonal complement $U^\perp$ is defined by the equations

$$\begin{aligned} -x_1 + x_3 &= 0 \\ x_2 + x_4 &= 0. \end{aligned}$$

Solution of Exercise 7.25: The points $l_1 := (0, 2, 1)$ and $l_2 := (2, 3, 0)$ lie in L, so that the direction vector of L, which is the difference of the two points, is $(2, 1, -1)$.

Likewise the points $m_1 := (1, 1, 2)$ and $(5, 3, 0)$ lie on M, so that its direction vector is $(4, 2, -2)$. This is a multiple of $(2, 1, 1)$, so the two direction vectors span the same (1-dimensional) subspace of $\mathbf{R}^3$, so the lines are parallel.

In order to compute the plane P containing M and L we use that P is uniquely determined (since $M \neq L$) by the condition that $l_1, l_2, m_1 \in P$. Thus

$$\begin{aligned} P &= l_1 + L(l_2 - l_1, m_1 - l_1) \\ &= (0, 2, 1) + L((2, 1, -1), (1, -1, 1)). \end{aligned}$$

In order to compute the cartesian form of P, we compute the orthogonal complement of $(2, 1, -1)$ and $(1, -1, 1)$: it is the 1-dimensional subspace of $\mathbf{R}^3$ spanned by a vector (a, b, c) satisfying $2a + b - c = 0$ and $a - b + c = 0$. This implies $a = 0$ and $b = c$, and c is a free variable. Thus the normal vector of the plane P is $(0, 1, 1)$ (or any nonzero multiple thereof). Taking into account that $(0, 2, 1) \in P$, we have

$$\begin{aligned} P &= \{v \in \mathbf{R}^3 | \langle v, (0, 1, 1)\rangle = \langle (0, 2, 1), (0, 1, 1)\rangle\} \\ &= \{v \in \mathbf{R}^3 | \langle v, (0, 1, 1)\rangle = 3\} \\ &= \{v = (x, y, z) | y + z = 3\}. \end{aligned}$$

For the second part, we first compute what a general point $q \in M$ looks like. From the above, we have

$$M = (1, 1, 2) + L(4, 2, -2),$$

so

$$q = (1 + 4r, 1 + 2r, 2 - 2r), r \in \mathbf{R}.$$

For $p = (0, 2, 1)$, this implies

$$p - q = (-4r - 1, -2r + 1, 2r - 1).$$

This vector needs to be orthogonal to the direction vector of L, which is $(2, 1, -1)$. Thus

$$\begin{aligned} 0 &= 2(-4r - 1) + 1(-2r + 1) - (2r - 1) \\ &= -12r \end{aligned}$$

so that $r = 0$, and $q = (1, 1, 2)$.

For the third part, we write X in cartesian form

$$X = \{v \in \mathbf{R}^3 | \langle v, a\rangle = b\}.$$

We need to compute the normal vector a and the number b. Since we seek $L \subset X$, the direction vector of L, $(2, 1, -1)$ needs to be orthogonal to $a = (\alpha, \beta, \gamma)$, i.e.

$$2\alpha + \beta - \gamma = 0.$$

Moreover, the vector $w := r - l_1 = (-1, 1, 0) - (0, 2, 1) = (-1, -1, -1)$ needs to be orthogonal to a as well, since r and $l_1 \in X$. Thus

$$-\alpha - \beta - \gamma = 0.$$

We solve this as $\alpha = 2\gamma$, $\beta = -3\gamma$, and $\gamma \in \mathbf{R}$ is a free variable, which we choose to be 1. Thus $a = (2, -3, 1)$. We compute b by observing that the equation defining X needs to be satisfied for r, so that $d = \langle a, r\rangle = -5$. Thus

$$X = \{(x, y, z) | 2x - 3y + z = -5\}.$$

For the last part, we have

$$N = r + L(w),$$

and we need to compute the direction vector w of the line N. Since $N \subset X$, we have $w \perp a = (2, -3, 1)$. Since $N \perp L$, we have $w \perp (2, 1, -1)$ (the direction vector of L). In other words, w is a vector in the orthogonal complement of these two vectors, or equivalently, the kernel of the matrix

$$\begin{pmatrix} 2 & -3 & 1 \\ 2 & 1 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 2 & -3 & 1 \\ 0 & 4 & -2 \end{pmatrix}.$$

Thus for $w = (\alpha, \beta, \gamma)$, γ is a free variable, which we can choose to be 2, so that $\beta = \frac{1}{2}\gamma = 1$, and $\alpha = \frac{3\beta - \gamma}{2} = \frac{3-2}{2} = \frac{1}{2}$. Thus

$$N = (-1, 1, 0) + L(\frac{1}{2}, 1, 2)$$

or, equivalently, for example:

$$N = (-1, 1, 0) + L(1, 2, 4).$$

Solution of Exercise 7.26: We form the matrix associated to the linear system and bring it into row echelon form

$$\begin{pmatrix} 1 & -2 & 0 & 1 \\ 0 & 3 & 1 & -2 \\ 3 & 0 & 2 & t \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -2 & 0 & 1 \\ 0 & 3 & 1 & -2 \\ 0 & 6 & 2 & t-3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -2 & 0 & 1 \\ 0 & 3 & 1 & -2 \\ 0 & 0 & 0 & t+1 \end{pmatrix}$$

This matrix has rank 2 if $t + 1 = 0$, i.e., if $t = -1$. Otherwise, for $t \neq -1$, its rank is 3. We compute a basis of U for $t = -1$ using Method 2.32. The result is $u_1 = (-2, -1, 3, 0)$ and $u_2 = (1, 2, 0, 3)$.

We apply Proposition 7.31:

$$\begin{aligned} w_1 &= u_1 / \|u_1\| = \frac{1}{\sqrt{14}} u_1. \\ w_2' &= u_2 - \langle u_2, w_1\rangle w_1 \\ &= (1, 2, 0, 3) + \frac{4}{14}(-2, -1, 3, 0) \\ &= (3/7, 12/7, 6/7, 3). \end{aligned}$$

The vectors w_1 and w_2' form an orthonormal basis of U. (It is possible, but here not asked for, to compute $w_2 = w_2'/\|w_2'\|$, then w_1 and w_2 form an orthonormal basis of U.)

The subspace $U^\perp$ consists of the vectors that are orthogonal on the basis vectors of U, i.e., the *kernel* of the matrix

$$\begin{pmatrix} -2 & -1 & 3 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix}.$$

A basis of this can be computed again by Method 2.32, for example $(2, -1, 1, 0)$ and $(1, -2, 0, 1)$ form a basis of $U^\perp$.

The subspace W will be the orthogonal complement of the line passing through v and w. The direction of that line is $v - w = (2, 0, 1, -3)$, so that

$$W = \{2x_1 + x_3 - 3x_4 = 0\}.$$

Solution of Exercise 7.27: We verify $B \in X$ by plugging in the coordinates of B:

$$2 \cdot 1 - 1 - 2 \cdot (-1) = 3,$$

so it satisfies the equation defining X. We now compute $C = p_X(A)$. The normal vector of the plane X is $n = (2, -1, -2)$. Thus $C = A - rn = (6 - 2r, -1 + r, -4 + 2r)$, for some $r \in \mathbf{R}$. In order for $C \in X$, this needs to satisfy the equation defining X, i.e.,

$$2(6 - 2r) - (-1 + r) - 2(-4 + 2r) = 3,$$

which simplifies to $21 - 9r = 3$, i.e., $r = 2$ and $C = (2, 1, 0)$.

The plane containing the points A, B, C is of the form

$$A + L(C - A, B - A).$$

In order to present this in cartesian form, we compute the orthogonal complement of $C - A = (-4, 2, 4)$ and $B - A = (-5, 2, 3)$, which is the line spanned by $w := (-1, -4, 1)$. The plane is then given by

$$\{x \in \mathbf{R}^3 | \langle x, w \rangle = \langle A, w \rangle\},$$

i.e.,

$$\{x \in \mathbf{R}^3 | -x_1 - 4x_2 + x_3 = -6\}.$$

The line that a) passes through B, b) is contained in the plane X and c) is orthogonal to the line passing through A and B, is given by

$$L = B + L(v).$$

By the conditions, $\langle v, B - A\rangle = 0$ and $\langle v, n\rangle = 0$. I.e., v is any non-zero vector in the orthogonal complement of $B - A = (-5, 2, 3)$ and $n = (2, -1, -2)$, for example $v = w = (-1, -4, 1)$. Thus

$$L = (1, 1, -1) + L(-1, -4, 1).$$

The direction vector of M_t can be found by choosing two points on M_t, for example $E = (0, 2, -1)$ and $F = (1, 2 + t, -2)$. So the direction vector of M_t is equal to $F - E = (1, t, -1)$. This needs to be orthogonal to $n = (2, -1, -2)$. I.e., $\langle n, F - E\rangle = 0$, which gives $2 - t + 2 = 0$, so $t = 4$.

References

[Bot21] F. Bottacin. *Algebra Lineare e Geometria*. Società Editrice Esculapio; English translation available in moodle, 2021. ISBN: 9791220877657. URL: https://books.google.it/books?id=efxVEAAAQBAJ.

[Nic95] W.K. Nicholson. *Linear Algebra with Applications*. Mathematics Series. PWS Publishing Company, 1995. ISBN: 9780534936662. URL: https://lyryx.com/linear-algebra-applications/.

[Oli11] Oswaldo Rio Branco de Oliveira. "The fundamental theorem of algebra: A most elementary proof". In: (2011). DOI: 10.4169/amer.math.monthly.119.09.753. eprint: arXiv:1109.1459.

Index